

1985

The development of a computer assisted assessment device for use in screening children referred for psychoeducational evaluations

Michael J. Buxton

College of William & Mary - School of Education

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Quantitative Psychology Commons](#)

Recommended Citation

Buxton, Michael J., "The development of a computer assisted assessment device for use in screening children referred for psychoeducational evaluations" (1985). *Dissertations, Theses, and Masters Projects*. Paper 1539618747.

<https://dx.doi.org/doi:10.25774/w4-p41y-wd94>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction:

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again, beginning below the first row and continuing on until complete.
4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases, the best available copy has been filmed.

**University
Microfilms
International**

300 N. Zeeb Road
Ann Arbor, MI 48106

Buxton, Michael J.

THE DEVELOPMENT OF A COMPUTER ASSISTED ASSESSMENT DEVICE
FOR USE IN SCREENING CHILDREN REFERRED FOR
PSYCHOEDUCATIONAL EVALUATIONS

The College of William and Mary in Virginia

Ed.D. 1985

University
Microfilms
International 300 N. Zeeb Road Ann Arbor MI 48106

Copyright 1986
by
Buxton, Michael J.
All Rights Reserved

THE DEVELOPMENT OF A
COMPUTER ASSISTED ASSESSMENT DEVICE
FOR USE IN SCREENING CHILDREN REFERRED
FOR PSYCHOEDUCATIONAL EVALUATIONS

A Dissertation

Presented to

The Faculty of the School of Education

The College of William and Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

BY

MICHAEL J. BUXTON

July 1985

THE DEVELOPMENT OF A COMPUTER ASSISTED ASSESSMENT DEVICE
FOR USE IN SCREENING CHILDREN REFERRED FOR
PSYCHOEDUCATIONAL EVALUATIONS


by

Michael J. Buxton

Approved July 1985 by


Charles Matthews, Ph.D.


Ruth Mulliken, Ph.D.
Ruth Mulliken, Ph.D.


Roger Ries, Ph.D.

Chairman of Doctoral
Committee

©1986

MICHAEL J. BUXTON

All Rights Reserved

DEDICATION

This work is dedicated to the two women of my life:

my mother June L. Buxton – who

continually encouraged and promoted my education

and to

my beautiful wife Linda B. Buxton – who

makes all things possible.

ACKNOWLEDGEMENTS

Major recognition is made of my wife Linda's incredible efforts. The very precise, detailed drawings of the 492 odd RSPM figures, very thorough proof-reading and her continued encouragement are greatly appreciated.

The tremendous assistance offered by my graduate committee, particularly my chairperson **Roger Ries**.

The personnel of Virginia Beach City Public Schools' support was essential and graciously provided. All parties involved deserve special thanks, particularly **Dr. Phillip Meekins**, the late **Dr. James Mounie** and **Dr. Andrew Carrington** who authorized the research. I plan to personally thank the Diagnostic and Counseling Staff whose participation was essential, as were the building principals who participated in the study, but feel that beach scenes and beer will offer a more appropriate setting.

Special mention of our programmer **David Rook** whose great talents transformed the very complex programming demands of the Macintosh into a very friendly, interactive assessment instrument.

Russell Pond who first demonstrated the Macintosh to me and provided considerable technical assistance throughout my work.

Enduring appreciation to **Dr. Margie Graf**, who first showed me the ropes of academia.

Linda, Mackie and I thank you all sincerely.

TABLE OF CONTENTS

	Page
DEDICATION -----	3
ACKNOWLEDGEMENTS-----	4
Chapter	
1. INTRODUCTION -----	7
Purpose and Importance of the Study-----	7
Statement of the Problem-----	11
Theoretical Rationale-----	14
Definition of Terms-----	19
Research Hypotheses-----	27
Sample descriptions & Data Gathering Procedures-----	29
Limitations of the Study-----	31
2. REVIEW OF RELATED LITERATURE-----	35
Efficacy of computer Assisted Assessment-----	35
Historical Review of Automated Testing Research----	43
Automated Projector Assessment Systems (APAS)---	43
Computer Assisted Assessment (CAA)-----	66
Response Time Research-----	79
Pertinent Research Regarding 'Motoric Efficiency'----	87
Research Regarding Potential of Macintosh Computer--	95
Summary of Research Review-----	100
3. METHODOLOGY-----	105
Population and Selection of the Sample-----	105
Treatment and Data Gathering Procedures-----	110
Instrumentation-----	118
Design-----	123
Specific Hypotheses-----	125
Statistical Analysis-----	127
Summary of Methodology-----	128

4. DEVELOPMENT OF THE AUTOMATED VERSION OF THE RAVEN STANDARD PROGRESSIVE MATRICES (RSPM)-----	130
Standardized Instruction for the Automated RSPM-----	158
Summary of Chapter Four-----	161
5. RESULTS-----	162
Hypothesis 1-----	163
Hypothesis 2-----	166
Hypothesis 3-----	168
Hypothesis 4-----	174
Hypothesis 5-----	180
Hypothesis 6-----	186
Hypothesis 7-----	189
Hypothesis 8-----	195
Hypothesis 9-----	204
 6. SUMMARY, CONCLUSIONS &IMPLICATIONS FOR FUTURE STUDY-----	 204
Summary-----	204
Conclusion-----	216
Implications for Further Study-----	219
APPENDIX A - RSPM Derived IQ Table-----	223
APPENDIX B - School Division Research Forms-----	228
APPENDIX C - Examples of MacPaint CAA RSPM Images-----	235
REFERENCES-----	237
VITA-----	247
ABSTRACT-----	249

CHAPTER I

INTRODUCTION

A. PURPOSE AND IMPORTANCE OF STUDY

Practicing school psychologists and Child Study Teams utilize a variety of screening devices to determine which of the many children who are referred to them will require complete psychoeducational batteries. The screening committee often bases this decision upon the child's apparent level of intellectual functioning which is believed to determine the student's potential to benefit from special instruction. This is particularly relevant when screening potentially Learning Disabled (LD) students who may be eligible for Special Education if it can be determined that they are not achieving at a level which is commensurate to their intellectual potential. During screening, assumptions about the child's intellectual potential are often based on psychometric procedures of limited reliability such as referring to the student's previous group administered SRA (STEA) scores, second

grade Kuhlman-Anderson IQ scores or through the administration of a brief intellectual screening device such as the Slosson Intelligence Test (SIT).

This study has developed an automated screening device which provided an indication of the child's fluid (Spearman's g), visually based intelligence, through the administration of the Raven Standard Progressive Matrices (RSPM). The use of a prototypic, automated assessment device offers the potential of providing screening committees with more thorough information regarding suspected Learning Disabled children prior to determining if the child should be referred for a complete psychoeducational battery. In addition, this device may ultimately provide practicing school psychologists with more complete information on which to base diagnostic decisions for those children who do receive complete psychoeducational evaluations. Such a device could help to screen larger numbers of potential LD candidates in a much more efficient and reliable manner while reducing the number of complete psychoeducational evaluations which are given to children who do not require special education services.

It is well known that individual psychometric assessment is a

time consuming procedure, requiring considerable effort from both the examiner and testee. A variety of strategies have been developed to make the task more efficient, usually by abbreviating the test or reducing the clinicians involvement in the administration (Denner, 1977). Abbreviation of a standardized test results in a final score that is based on less information than the original instrument required. Thus the reliability and validity of the test suffer, often to an undetermined degree (Calvert and Waterfall, 1982). One means of reducing the amount of examiner time required for direct one-on-one testing without reducing the amount of information obtained is through the use of computer assisted assessment (CAA).

The application of a microcomputer for the administration of the Raven Progressive Matrices has been reported by one researcher (Rock & Nolen, 1982). In this work, which used the Coloured Progressive Matrices (CPM), no significant difference was found between the machine administered version and the manually administered test. In addition, the results were correlated with the results of WISC-R. No significant differences were found between the two different CPM forms of administration. The article suggests that

generalization of these findings is limited due to "the relatively small number of subjects (N=15) and the lack of rigorous experimental controls". The research considered here attempts to extend and enhance the findings of the aforementioned study through the use of more sophisticated hardware, a larger sample of subjects, and the use of two additional computer obtained psychometric variables: item and resultant cumulative response time measured in milliseconds; and a measure of 'motoric efficiency'.

This investigation is the first known study to employ CAA techniques in comparing the manual versus automated versions of the RSPM in a counter-balanced, test-retest design. In addition, the study made a post hoc comparison of the effectiveness of the CAA version of the RSPM in predicting WISC-R verbal, performance, full scale, and subtest scores as well as various WISC-R factors derived through factor analysis, e.g. Kaufman's Freedom From Distractibility Factor. (1975).

B. STATEMENT OF THE PROBLEM

This study consisted of two parts. The first part employed a counter-balanced, test-retest design comparing the CAA version of the RSPM developed in this study with a manual paper and pencil administration of the same instrument. The experimenter attempted to obtain a sample population of sixty children ($N = 60$) in grades 4 through 6 for inclusion in this portion of the experiment. This sample size allowed for groups of $N = 20$. Such a sample size should offer relatively potent comparisons, without unduly disturbing a great number of childrens' educational programs.

The rate of subject attrition was overestimated, consequently the sample sizes in each group were slightly larger than had been projected. The groups in this portion of the study were used to assess the test-retest reliability of the CAA version of the RSPM with itself (CAA - CAA, $N=21$); and the order effect of the CAA version first followed by the manual version (CAA - Manual, $N=21$); and the reversed order (Manual - CAA, $N=23$). No test-retest comparison of the manual versus manual portion of the RSPM was

made, as the reliability and validity of the manual version is well researched and quite acceptable (Raven, 1983).

The second portion of the study involved the post-hoc comparison of previously obtained (within twelve months of the study period) WISC-R scores with that of the CAA administered version of the RSPM. As previously stated, the post-hoc analysis examined the effectiveness of the CAA version of the RSPM in predicting WISC-R verbal, performance, full scale, and subtest scores, and significant WISC-R subtest factors previously derived by researchers through factor analysis. The study compared the RSPM IQ score, which was developed using the "Smoothed Summary Norms (1979)", which were "based on a nationally representative sample of British schoolchildren, excluding those attending special schools" (Raven, 1983, p. SPM21). These norms provide percentile scores, which were converted into standard scores in this study to allow direct linear comparisons with the previously administered WISC-R scores.

The 1979 British norms have been found quite applicable to children in the United States as the result of a recently completed standardization of the RSPM in the United States (Raven, 1985).

The applicability of the standardized norms for the automated RSPM was also investigated as part of the first portion of this study. Statistical tests indicated that the automated RSPM was quite reliable and demonstrated a high degree of comparability with the manual version of the instrument.

In addition to comparing previously derived and researched scores (eg. RSPM Percentile Scores) this research also examined the relationships of five newly developed CAA obtained parameters. Two measures of 'motoric efficiency' (cumulative initial phase; and cumulative secondary) and four measures of response time (cumulative initial, pre-response time; cumulative intermediate response time; and cumulative final response time). In addition, total test time (real clock time); was also examined in relation to the other psychometric parameters.

C. THEORETICAL RATIONALE

Miller (1968) asserted that automated testing offers the potential for psychologists to obtain equivalent or greater amounts of psychometric data about testees as manual testing without spending the considerable amounts of time generally required to perform manual tests. Research has shown that in counter-balanced, test-retest studies, comparing standardized manually administered tests with automated versions, there are generally no significant differences between the two versions of the test (Elwood, 1969, 1972a, 1972b; Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982).

Past research on automated testing has been flawed through the relatively limited capabilities of the equipment used, failure of this equipment to provide an equivalent level of interaction between the testee and the machine as between the testee and an examiner, and lack of researchers' compliance with the procedural guidelines for administration established by the developers of the various test instruments assessed. These factors have provided numerous

indications that the automated versions of the tests are somewhat more difficult (generally to a non-significant degree) than the manual versions.

The majority of the previous automated testing research has been performed on sample populations of relatively small numbers of subjects or with subjects of limited intellectual capabilities. Very little work has been done with children of normal intellectual ability. Thus, additional research investigating the efficacy of CAA in evaluating children is of great interest and importance to practicing school psychologists.

This study developed a diagnostic screening device designed to utilize CAA techniques, using hardware with which individuals could interact quite easily. This study suggested that a major reason for the generally observed difficulties of previous automated testing research stemmed from the failure of the equipment to provide the testee with as high a degree of feedback from their responses as that provided by the manual versions of the tests being evaluated. This lack of feedback appeared to reduce the testee's opportunities to learn from their previous responses or modify them. In addition, the automated

versions of the tests, at times, also failed to comply with the guidelines for the manual administration of the test, thereby putatively further increasing the difficulty of the test. In fact, there is no known research on the RSPM which has complied with Raven's, et. al. (1977, 1978; Watts, Baddeley, & Williams, 1982) directions for administration of his test in an automated format. The research considered here attempted to comply fully with Raven's recommendations for automation in addition to having the CAA version comply as closely as possible to the guidelines for manual presentation of the RSPM.

It is well known that many intellectual tests utilize measures of response time in the generation of their test findings. This study introduced the use of very precise response timing techniques to the RSPM. The use of a three-phase response timing allows the examiner to fragment the testee's individual responses thereby providing an opportunity to analyze and make hypotheses about the problem-solving and personality factors (eg. extroversion, compulsivity, tendency to daydream) of the testee, and these subsequently derived factors' relationships to other psychometric parameters. In addition, the

measurement of response time as a form of more precisely analyzing testees behaviors may ultimately increase our understanding of the relationship between directly measurable cognitive factors, such as response time relationship with the more global, psychometrically derived concept of intelligence.

A major concern of theorists working in the area of automated testing is that the subjective, qualitative elements of the testee's response are lost through automation, just as they are in group testing procedures. Information regarding the subject's inferred level of motivation, how diligently he approached the task, his tolerance to frustration when experiencing personally difficult items and the purposefulness of his activity in working with the test materials are all factors of considerable interest to psychologists. Unfortunately, a simple machine administered and scored test will only reflect the effects of these factors. Consequently, it has been purported that the validity of a particular score cannot generally be determined from CAA test results, while such considerations are an important element of manual administration techniques. However, unlike group testing procedures, CAA has the potential of objectively measuring numerous

individually based variables related to test behavior which can help to determine the validity of the obtained scores.

Several means of obtaining an indication of the subject's purposefulness in performing the CAA version of the RSPM was developed. A measure of 'motoric efficiency' attempted to quantify the heretofore subjectively based psychometric factors concerning the testee's behavior during test administration. The use of a two-phased measure of 'motoric efficiency' provides the examiner with information regarding the initiation, frequency and extent of non-purposeful behavior occurring on the test. The use of initial, intermediate and final response times also provides the examiner with indications of the degree of on-task behavior the testee demonstrated in working on the automated RSPM.

Consideration of the relationships between these measures of problem-solving style and response time, which are behaviorally based measures, and the psychometrically derived scores on the RSPM and WISC-R will be examined in the forthcoming discussion of the results of this study.

D. DEFINITION OF TERMS

The terminology used in this study is presented below in alphabetical order.

Errors: Four categories of errors were recorded by the CAA device which was developed as part of this study. They were:

1. **Fatal:** Scored errors which the subjects clicked as their response and subsequently "locked in". The actual position of the subject's choice was recorded to allow the determination of any response set which may have adversely affected the subjects' performance.
2. **Non-Fatal:** This class of errors provided an indication of the degree of certainty that the testee exhibited in his responses. A non-fatal error was any answer that the subject chose and subsequently changed. Whether the item was incorrect or correct, or whether the subject eventually obtained a correct answer on the item did not modify the recordation of this variable. Clicking the mouse on areas of the screen which did not contain an answer, was also scored as a non-fatal error.
3. **Cumulative Total of Subjects' Non-Fatal Errors:** The total number of incorrect responses that the subject made prior to subsequently changing them into the correct answers.

4. **Total Number of Correct Responses:** This score, which was computed by taking the total number of items (60) less the number of fatal errors, was the parameter used to make comparisons with other RSPM research and to derive IQ scores.

Motoric Efficiency: This was a linear measure of the subject's movement of the cursor across computer screen. The computer was programmed to count the number of pixels (dots on the screen which produce the Images) that the cursor passes through in its movement. This newly derived measure was possible to obtain due to the "bit-mapping" format of the Macintosh computer. The high resolution of this computer allowed the development of this linear measure to provide accurate indications of subject's mouse movements in excess of 1/80 of an inch. This research obtained two measures of this behavior and subsequently derived four psychometric parameters from these measures. They were:

1. **Initial Phase of Motoric Efficiency:** This was a measure of the linear movement of the cursor from the initial placement as the screen refreshed itself with the next RSPM image until the subject made his initial answer by clicking the mouse when the

cursor was on the item of his choice. Such a measure was obtained and stored for each scorable item.

2. **Second Phase of Motoric Efficiency:** This was a measure of the cursor's movement, starting from the when the subject first clicked his tentative answer until he "locked in" his response. This parameter was obtained for each scorable item. This measure also determined the extent of cursor movement during the subject's process of making non-fatal errors (see fatal and non-fatal errors).
3. **Cumulative Measure of Initial and Secondary Phase Motoric Efficiency:** The respective results of each items' initial and secondary motoric efficiency scores was summed. The statistical manipulation of these parameters was performed using the cumulative parameters.

Mouse: This device was a small hand-held object which has a single button on its top surface which could be depressed or "clicked". A ball on the bottom of the mouse allowed it to roll on the table, and through an optical-electronic process the mouse's movement was depicted on the computer screen by the movement of a cursor or arrow. The mouse was connected to the computer screen by a fine insulated wire which gave it the appearance of an actual mouse (palm-sized body, long thin tail).

Through the use of the mouse, the subject executed his choices by moving the cursor and then clicking the mouse's button. Feedback was provided to the subject through highlighting on the screen.

1. **Clicking:** The subject interacted with the computer by moving the mouse about the table, thereby moving the cursor on the computer screen. For example, upward movement of the mouse provided corresponding upward movement of the cursor on the computer screen. The subject could use the mouse to move the cursor to any area of the screen. When the cursor was positioned over the response of the subject's choice, he depressed the mouse's button (which produced a clicking sound) to indicate his choice. Subjects were required to click the mouse in order to chose their answers. In addition, subjects were able to change their answers by moving the cursor to a different answer and clicking the mouse button on the newly selected answer. This item was then subsequently highlighted. Once the subject was satisfied with his answer, he was required to "lock in" his response by moving the cursor to a designated area of the screen and clicking the mouse button again. The CAA program would then present the next test item.
2. **Highlighting:** When the mouse button was clicked on an appropriate sensitized area, the background of the area was

programmed to change from white to another shade thereby providing feedback to the subject that his response had had an effect on the computer. Within this study, the highlighting format of the Macintosh computer was used to produce a black rectangle around the subject's choice, thereby providing feedback. If the subject decided to change his answer, his new choice would become highlighted. Only the most recently selected item would be highlighted on the computer screen. Highlighting did not occur if the subject clicked on an inappropriate area of the screen, thus providing feedback that an inappropriate response had been made. Several other test stimuli conditions were developed to reduce the possibility of subjects making inappropriate responses (see prompt changes and programmed prompts).

Prompt Changes: The Macintosh programming format allowed for the displays on the screen to change in accordance with the task that the subjects should perform. In the newly developed automated version of the RSPM the area in the upper portion of the computer screen was designed to indicate to the subject what task they were expected to perform. Upon initial presentation of an item a prompt stating "Move your arrow to the correct answer and click" was presented at the top portion of the screen. When the subject chose an answer and clicked on it, this prompt changed to "Check or change your

answer and lock it in here->". A box to the right of this prompt was programmed to be sensitized so that when the mouse-controlled arrow was placed inside the box and clicked, the subject's choice was 'locked in', recorded and the next item of the test was shown.

Programmed Prompts: Should a subject fail to appropriately interact with the CAA device, prompts generated by the computer could be used to provide assistance. The need for many programmed prompts was not anticipated in this study. However the potential for their use on more complex forms of CAA instruments seems quite promising. This study's research proposal anticipated the possible need for a programmed prompt if a subject tried to 'lock in' his response prior to choosing an answer. A programmed prompt or dialogue box which would provide a window that overprinted the screen and stated: 'Please choose and click your answer, before you lock it in' was proposed. Preliminary field testing of the RSPM CAA instrument indicated no need for such a programmed intervention, as the

subject's readily grasped the two component response format of the test.

Response Time: Four measures of response time were taken in this study. From these, five CAA derived psychometric parameters were derived. These were:

1. **Total Test Time:** This measure noted the starting time of the test from the computer's internal real time (AM/PM) clock at the beginning of the test (during presentation of the initial practice items) and again at the completion of the final test item. The total clock time spent on the entire CAA test procedure is recorded as part of the subject's CAA test scores.
2. **Initial Response Time:** This measure recorded the time from the initial presentation of each item, (not counting screen refreshment) until the subject moved the mouse out of the assigned area where he had locked in his previous response (upper right hand side of the screen). This movement was elicited by the prompt 'Move your arrow to the correct answer and click'
3. **Intermediate Response Time:** This measure recorded the elapsed time starting when the subject first moved the mouse out of the designated area in the upper right hand portion of

the screen area until they clicked their initial selection, thereby highlighting their tentatively chosen answer.

4. **Final Response Time:** This measure recorded the time from the subjects' initial answer selection until they 'locked in' their answer by clicking the mouse in the designated area located in the upper right hand area of the screen. This area was clearly designated by the prompt statement: 'check or change your answer and lock it in here'. The subject was then presented with the next item of the RSPM and the response time measures continued as defined above for each item.
5. **Cumulative Response Times:** The initial, intermediate and final response times for each item were totaled to yield cumulative measures of each parameter. The statistical manipulation of response time parameters was performed using these cumulative times.

Screen Refreshment: After an item was chosen for each item and 'locked in' by the subject, the next item was shown on the screen. A time period of approximately one second in the newly developed CAA device was required for the computer to search its hard-disk drive for the next item, read it off the disk, decode the compressed storage version of the item (unpacking) and depict it on the screen.

E. RESEARCH QUESTIONS

This study was designed to attempt to answer the following questions:

1. Will there be a non-significant difference between the computer assisted administration of the RSPM and the manual administration of the the same test when compared in a counter-balanced, test-retest design?
2. Will the computer assisted administration of the RSPM demonstrate a statistically acceptable degree of reliability and concurrent validity when measured in a test-retest format?
3. Will the computer assisted administration of the RSPM show any significant order effect (CAA - then - Manual versus Manual - then - CAA) in appropriate statistical comparisons?
4. Will an automated (computerized) administration of the Raven Standard Progressive Matrices (RSPM) demonstrate a statistically significant relationship with the WISC-R when compared using a post hoc analysis?

5. Will the automated RSPM yield a degree of concurrence with the WISC-R scores comparable to previously established levels of concurrence reported in test literature for the manual administration of this instrument?
6. Will the results of the automated testing procedures yield any significant inter-relationships with the results of previously administered WISC-R subtest scores, or Verbal, Performance, and Full Scale scores when analyzed using the appropriate statistical procedures?
7. Will the results of the automated testing procedures yield any significant inter-relationships with the previously derived WISC-R subtest factors derived through factor analysis.
8. Will the unique psychometric parameters of subjects' three-phased item and total response times and two-phased measure of 'motoric efficiency' provide any statistically significant quantitative interrelationships when compared with other group and individual data obtained through automated and/or previously administered manually obtained psychometric data?

E. SAMPLE DESCRIPTIONS AND GENERAL DATA GATHERING
(directional) PROCEDURES AND/OR INTERVENTIONS.

This study employed an Apple Macintosh 512 kilobyte computer to administer the automated version of the RSPM. The eighty-five (85) subjects were upper elementary students in the 4th, 5th and 6th grades who were randomly chosen for participation in the two different components of the evaluation. The first phase of the research (test-retest validation study of the CAA vs Manual versions of the RSPM) employed students who were chosen from grade level class lists of two elementary schools. The researcher randomly selected ten subjects from each grade level in each of the two schools for administration of the test-retest portion of the proposed research.

The sixty-five (65) subjects in the initial portion of the research were administered the paper and pencil version and the Automated version of the RSPM in a test-retest counter-balanced design format which considered the order effect of the following groups: CAA first- Manual second (N=21); Manual first- CAA second (N=23); and CAA first- CAA second (N=21) to

determine the reliability of the CAA version of the RSPM. No measure of the test-retest reliability of the manual version of the RSPM was made, as the validity and reliability of this instrument has been well researched and documented. (In the research literature) A two week interval between the test and retest portion of the research was provided to diminish practice effect, which was controlled, in part, by the counter-balanced nature of the design.

During the second phase of the study a post hoc comparison of the results of the CAA version of the RSPM with the results of previously administered WISC-R evaluations was performed. Subjects were again 4th, 5th and 6th grade students (N = 20) who were randomly chosen from a list of students provided by the Director of Pupil Personnel. Subjects included in this portion of the study had initially been referred for a psychoeducational evaluation due to suspected learning disabilities. Subjects were randomly selected from those who had been administered the WISC-R by a Virginia certified school psychologist within the previous twelve months prior to the implementation of the study. Each subject's WISC-R subtest,

Verbal, Performance and Full Scale scores were compared with the various CAA obtained scores.

A standardized set of instructions was provided to the CAA subjects throughout the study. The experimenter of this study and the ten other certified school psychologists who participated in the second phase of the study used these directions. They are shown verbatim in Chapter Four - Development of the Automated Version of the Raven Standard Progressive Matrices – section of this study.

G. Limitations of the Study

The proposed study did not provide an indication of the subjects' perceptual skill development. The age of the subjects selected for this study (approximately 8 - 12 years of age) was made in part, due to the well known finding that children generally develop compensatory skills by age 8 1/2, which diminish the adverse effect of perceptual difficulties upon academic progress. Nevertheless, perceptual skill development remains an uncontrolled variable in this study.

This study will only assess one component of a subject's psychological composition. This is due to the limitations of current micro-computer hardware. The inability to obtain a reliable and portable memory storage device (hard-disc of 10 Megabyte capacity or more) during the initial planning portion of this study forced the developer of this research to limit the focus of this study to the development of a screening device which only examined intellectual potential. Future research will hopefully investigate the development of perceptual, academic and projective devices to provide for more thorough,

comprehensive CAA forms of assessment. (See Chapter Six - Summary, Conclusions and Implications for Further Study)

The generalizability of this study to populations beyond the age range assessed is questionable, due to numerous developmental factors which may affect the capabilities of CAA with older or younger subjects. Further work in delineating the relationship of CAA with different populations would appear advisable.

The generalizability of this study's findings to predict the efficacy of CAA techniques with other psychometric devices is also questionable. The RSPM is very well suited to the CAA format which was used in this study. The application of this study's newly developed techniques, parameters, and equipment to other test applications should be performed cautiously and only on a prototypic basis.

Although a premise of this study has been that the use of CAA devices will enhance psychologists' productivity, the use of CAA instruments must be qualified. CAA instruments must exclusively be used under the supervision of trained psychologists. The use of CAA screening devices results must only be interpreted by competent

psychologists, trained in the statistical and psychometric principles fundamental to the interpretation of psychological data.

CHAPTER II

REVIEW OF RELATED LITERATURE

A. EFFICACY OF COMPUTER ASSISTED ASSESSMENT:

Miller (1968) made a thorough argument for the use of automated psychological tests. He emphasized that the potential of automated testing to offer a higher degree of standardization of test administration and more precise recordation of test responses would allow for higher reliability of the test instrument. Miller also asserted that the psychologist could be relieved from the mechanical skills required in testing, thereby increasing their time for performing

tasks which could not be performed by machine. Miller proposed that automated testing could also require less of the examinees time. However, he qualified that the relative efficiency of the machine would be dependent upon the type of person assessed.

Miller also suggested that highly trained personnel would no longer be required for the administration of tests, thus increasing the amount of time the trained psychologist would have for test interpretation and clinical work. As a result of all these factors, Miller asserted the cost of automated test presentation would compare favorably with conventional, manual administration techniques. Elwood (1969) also pointed out the potential time and money savings offered by automated testing in his initial work on an automated WAIS administration. Elwood's later (1972a) work actually quantified the savings of automated testing and found that automated testing could be performed at less than half the cost of face to face manually administered tests.

Elithorn, Mornington and Stavrou (1982) asserted that:

"the majority of the tests in clinical use involve highly trained administrators in the expenditure of considerable amounts of time and effort for relatively low returns in

terms of relevant useful information. Automated testing in its simplest form can be justified in terms of manpower savings alone." (p.247-248)

Colles (1975) stated that automated testing procedures should provide the maximum information from an existing manual test with the minimum expenditure of the clinician's time.

Watts (1982) discussed the additional advantages offered by automated testing. She indicated that automation allows for increased control over presentation conditions. Response times could be more precisely measured. She also concurred with the previously mentioned researchers' findings that automated testing, more importantly, does not require a highly trained psychologist for administration and therefore "a potentially far greater number of subjects can be tested" (p.332). Thus, differential examiner effects can effectively be diminished while obtaining precise behavioral measures of the subjects' response times. Another apparent advantage of computer monitored response times is that they are obtained in a covert manner which probably would not produce performance anxiety to the same degree as an examiner with a stop watch.

Several researchers have reported that testees prefer to work with automated systems rather than engage in face to face test sessions. Evans (1975) and Calvert and Waterfall (1982) report that shy individuals appear to work more at ease with the computer and other testees appeared to simply be fascinated by interacting with the machine. Watts (1982) reported that in her work with geriatric subjects, she encountered three testees who refused to complete the test. All three individuals were working on the paper and pencil version of the instrument being assessed, when they refused to proceed further on the test.

Stallard (1982) proposed that the adoption of automated testing procedures would offer considerable gains to handicapped individuals. He asserted that computers would allow the schools to enhance their ability to monitor the educational process of exceptional children. In addition, he stated that the computers' ability to increase psychologists' productivity would increase the availability of diagnostic and testing services to children.

Weinman (1982) suggested that the further development and utilization of computer assisted assessment would allow psychologists

increased understanding of the ultimate psychometric question: What is it (cognitive factors, neurological predisposition, personality factors, intelligence, etc.) that makes people score at a particular level on a test? Weinman indicated that this level of inquiry is quite an advance beyond traditional "psychometric tests [which] have been designed to provide accurate ways of classifying individuals in terms of specific abilities. The information obtained from these tests provide very little or often no indication about the cognitive processes which an individual employs in attempting a test" (p.321). Through enhanced understanding of the underlying cognitive processes used by people of differing ability levels, which heretofore has generally only been explored by experimental psychologists (not psychometricians), psychologists may develop more extensive models of these cognitive processes.

The use of empirically based (using cognitive information previously unattainable without computer mediated testing) cognitive models will allow for further refinement of the measurement of these models. The potential exists for therapeutic interventions, both educational and emotional, to utilize the results from the

measurement of these cognitive processes and develop remedial processes designed to modify specific cognitive deficiencies when they are identified. The effectiveness of such intervention strategies could also be measured on an ongoing basis, thus allowing modification of the treatment approach as part of a routine process of assisting the client to more effectively interact within their world.

The above brief discussion on the efficacy of automated testing has not explored the issue regarding the relative degree of reliability and concurrent validity of automated tests when compared with the manual tests on which they are based. This will be addressed in forthcoming sections of this literature review.

In summary, researchers have found automated tests to be appealing to testees, to an equal or greater degree than manual tests. They provide the opportunity to obtain psychometric parameters unattainable without the use of highly mechanized apparatus. Automation offers the potential for higher levels of reliability due to reduced 'examiner effects', a higher level of test administration standardization, and more precise measurement of response time without the apparent induction of heightened performance anxiety that

is generally the result of an examiner closely timing a subject's performance. Automated testing has also been found to enhance psychologists productivity, allowing them increased time to perform dyadic or group interpersonal tasks with clients, which cannot be performed by a machine. This enhanced productivity also represents a significant cost savings, which could allow for more extensive diagnostic testing, thereby possibly producing more accurate therapeutic decisions.

Most importantly, this initial section suggests that application of automated testing techniques, and the heretofore unattainable (discussed in forthcoming sections) psychometric parameters such as precise, multi-factored response times and 'motoric efficiency' measures, offer psychological measurement a richer understanding of mental processes.

Psychometrics, behaviorism and neuropsychology may thus integrate their areas of exploration. The potential exists for psychology to move away from the very restricted use of tests as indexes of relative (normative) ability and instead make it possible to determine what an individual can do and how he does it. The

answer to these questions may allow psychologists to develop strategies to assist less able individuals to adopt more effective cognitive strategies. The resultant focus on cognitive deficits (measured by direct behavioral assessment) and means for remediating them would appear to provide a more efficient and beneficial framework to address individual differences than the current concern for an individual's placement within the normative distribution of an abstract (possibly artificial) psychometric parameter.

B. HISTORICAL REVIEW OF AUTOMATED TESTING RESEARCH:

Throughout the above discussion the term automated testing was used, rather than the term computer assisted assessment which is contained within the title of this research proposal. Automated testing is a broader term and is an area in which considerable work has been performed, prior to the recent development of inexpensive and relatively powerful microcomputers. Thus for purposes of consistency, the term automated testing will continue to be used as a general term which may include both Computer Assisted Assessment (CAA) and Automated Projector Assessment Systems (APAS).

1. Automated Projector Assessment Systems (APAS)

Automated Projector Assessment Systems (APAS) appear to have been first employed by Gedye and his coworkers in 1966. (Gedye, 1966, 1967, 1968; Gedye & Gaines, 1967; Gedye & Miller, 1969, 1970; and Gedye and Wedgewood, 1966) They developed their assessment

device from a teaching machine. The machine, which was referred to as a ts-512, was used to present a Pictorial Paired Associate Learning Task (PPA) to a variety of subject populations. The actual task assessed by the PPA required the selection of the odd (unrelated) item from an array of three items. The initial items required only perceptual differentiation but gradually included increasingly more difficult conceptual items. The responses and latencies between responses were printed out by an electro-mechanical system. The ts-512 proved to be a sensitive and reliable technique for assessing the effects of therapy (both drug and rehabilitation) on intellectual and cognitive disorders.

Levy and Post (1975) explored the feasibility of using the ts-512 presentation of the PPA in the assessment of psychogeriatric patients' cognitive functioning. They reported that the technique was easy to administer and was more readily accepted by patients than standard psychological tests. The authors indicated that the machine's recording of results in a format ready for analysis was very useful. The test-retest reliability was found to be "reasonably high" ($r=0.58$), although not as high as desired. The machine

administered PPA was found to concur quite closely with the Subjects' performance on standard psychological tests.

Exton-Smith (1980) used the automated PPA with 30 elderly subjects who were also evaluated by a battery of standard tests. He reported high correlations between the PPA and other assessments. He found that the automated procedure afforded a global assessment of intellectual functioning and day to day coping capability.

Flowers (1968) developed an automated short-term visual memory test which required subjects to indicate repeated pictorial items. He found the test effective in assessing the memory losses occurring during the post-traumatic period of concussions.

Elwood (1969, 1972a, 1972b) created an automated version of the Wechsler Adult Intelligence Test (WAIS). As previously mentioned, he found that administration costs were less than half that of conventional administration techniques. In his test-retest comparisons between manual and automated WAIS administrations, he generally found good agreement between the two methods. Elwood's 1972b study examined the validity coefficients of automated vs. manual WAIS subtest and scale scores and assessed the order effect

of which version of the test was administered first on validity coefficients and the manner in which these gains occurred as a result of order of version of test administered (i.e. practice effect). The results revealed that analysis of counter-balanced designs needs to be carefully considered.

Elwood (1972b) found the automated WAIS first group (N=20) obtained scores that did not significantly differ from the validity coefficients of the manual version first group (N=20). However examination of the mean scores of the respective groups demonstrate that the magnitude of the practice effect resulting from the second administration of the WAIS within an approximate 10 day period, favored the group that was offered the manual version first. When Elwood (1972b) analyzed the mean score increases between the two groups, he found significant differences favoring the manually administered first group for the vocabulary, digit symbol, object assembly, performance IQ, and full scale IQ. Elwood accounts for this difference as a result of the very adverse effect of the automated version of the digit symbol and object assembly subtests which used the time scoring norms from the manual version of the WAIS. These

low automated performance subtests subsequently lowered the performance and full scale IQ scores which are derived from these subtest scores. Elwood failed to account for the depressed vocabulary score.

A possible explanation for the lower scores on the WAIS subtest scores of subjects who were first administered the automated version is that subjects failed to gain the same experience from the automated version's manner of administration as those subjects who on their first test interacted with an examiner. The vocabulary subtest, with the opportunity for the examiner to ask the testee to "tell me more about it" (Wechsler, 1981, p.54) may have enhanced the subjects ability on the automated retest to effectively dictate into the automated version's microphone the more complete answers derived from the initial manually administrated testing format.

Elwood failed to address the effects of automation on making the object assembly and digit symbol subtests more difficult than the manually administered versions of these tests. The relatively simpler presentation of the manually administered object assembly

and digit symbol subtests (puzzles on a table and a pencil & paper task) appear to have assisted those subjects who experienced them first (manually). The automated first group, in order to complete the object assembly subtest, had to press a ready button, look inside a drawer that contained the standardized pre-disarranged pieces, assemble them and close the drawer. On the automated digit symbol subtest, which produced a extremely significant lower order effect score ($p=.001$), Subjects were required to use a novel set of buttons to make their responses. Both of these automated tasks seem to have induced more difficulty than the original manual tasks.

These findings stress the importance of minimizing changes to tasks found on pre-existing tests and the importance of obtaining validity coefficients between manual and automated tests. In addition, it is suggested that the automation of pre-existing manual tests should be done very selectively, to minimize this effect, or the tests should be independently re-normed. Due to this factor, a counter-balanced test-retest comparison of the manual vs. automated version of the Raven Standard Progressive Matrices (RSPM) test was included in the study. The research findings to be discussed below

clearly indicated that automation of the RSPM did not induce the pronounced change in the level of difficulty that appears to have adversely effected Elwood's work with the WAIS.

Although not discussed in Elwood's research reports, the effect of subjects' perception of isolation stemming from the automation of test items originally designed for dyadic interaction between an examiner and the testee appears to be factor in his research. His description of a testee pulling open drawers containing the performance items and speaking verbal responses into a microphone for later transcription and scoring, would appear to somewhat change the subjective nature of the test.

A similar finding was made by Brierley (1971). He presented an automated projector version of Anstey's dominoes, a test designed as a correlate of the Raven's progressive matrices. He examined the test-retest reliability and validity of the automated version of this test with the manual version and found that little was lost concerning these psychometric factors when subjects were of at least average ability. Brierley also reported that the automated procedure enhanced subject's motivation. However, some subjects

reported the relative isolation and lack of feedback unpleasant. Thus, the lack of interaction in the narrowly programmed Automated Projector Assessment Systems (APAS) is emphasized. Brierley comments that the isolation of subjects gave the examiner limited opportunity to observe the subject's responses, a psychometric parameter of considerable importance to the clinical psychologist. This factor is another limitation of the APAS format of automated testing, wherein the machine is only able to record "a simple operant response", such as a button push which thereby produces very limited feedback to the subjects, and provided limited data to the examiner. (Overton & Scott, 1972, p.642)

The importance of explicit directions, which may require modifications to accommodate the automated testing format, was emphasized by Overton & Scott. They found a non-significant, detrimental effect in their counter-balanced comparison of the automated vs. manual presentation of the Peabody Picture Vocabulary Test (PPVT) when the automated version was given first. Their subjects were retarded adults (approximate average age 24, approximate mean IQ = 50). As an experimental condition, only those

subjects who reached a basal were included in the study. The authors found a significantly ($p < .01$) greater failure rate among those who received the automated test first, in addition to the aforementioned non-significant mean score difference. The authors assert that "the disadvantage of the automated version could probably be eliminated if the quite minimal pretraining were extended". (p.6)

Knights, Richardson & McNarry (1973) found a similar detrimental effect in their comparison of manual vs. automated versions of the PPVT and Raven Coloured Progressive Matrices (CPM) when the automated version was administered first. Their subjects were retarded children (mean age 15.7; mean IQ = 43) who were administered the PPVT and then the CPM. One-half of the children were administered the automated versions of the tests first and subsequently offered the manual versions two weeks later. The other half received the tests in the opposite order. As in other test-retest studies, there was not a significant difference between the test format (manual or automated), yet the second version of the tests yielded higher scores. There was, however, a significant difference when the first testing was the automated version of both the PPVT

and CPM. Those children who were first tested with the automated version obtained the lowest score and made the greatest improvement when offered the manual version during the retest phase of the experiment.

Knights, et. al. (1973) asserted that the significant difference stemming from order of administration may have resulted from the failure of a number of children to understand the initial demonstration item on the automated version of the CPM. They emphasized that in future studies, they would modify the first CPM item as a number of the children had difficulty in understanding the demonstration item provided. These researchers also used this factor to account for a lower test-retest correlation of the CPM (which was nevertheless favorable) compared with that obtained by the PPVT.

Thus, both Knights, et. al. and Overton & Scott stress the importance of providing explicit directions which are appropriate for the population of testees. In addition, they may have needed to consider Elwood's assertion that automated testing was adequate for subjects of at least average intelligence. Also, a graphic indication of the subject's responses to enhance their perception of feedback

may have overcome Brierley's (1971) previously noted observation that some subjects found the lack of feedback unpleasant in APAS testing. The use of Computer Assisted Assessment (CAA) equipment, rather than the Automated Projector Assessment Systems (APAS) utilized in the above research would allow for more feedback and more thorough interaction between the subjects and the machine during the introductory portion of the evaluation to ensure the subjects' understanding of the test items. Seemingly, the greater capability of the CAA format of automated testing will overcome many of the limitations of automated testing discussed by Elwood (1969, 1972a, 1972b); Overton & Scott (1972); and Knights, et. al. (1973).

The immediately preceeding studies' findings demonstrated adequate test-retest validity coefficients in all cases. However, in all cases they failed to favorably compare with manual test administration due to possible increased item difficulty stemming from automation and/or lack of understanding of directions by subjects. Specifically, Elwood (1972b) and Knight et al. (1973) showed a significantly greater degree of practice effect when the

apparently more difficult automated version was presented first (subjects then showed greater gains on the apparently less difficult manual versions). Overton and Scott, (1973) with their very large sample size (automated or manual group N=120) found no significant difference between order of administration. They were not concerned regarding practice effect as their research design alternated forms PPVT-A or PPVT-B in their test-retest design. Their findings with institutionalized mentally handicapped subjects did suggest that the automated version was more difficult (due to inadequate directions or pre-training) since a significantly greater number of subjects failed to achieve a basal level when administered the automated test compared to the manual version of the test.

This study attempted to avoid the difficulties that have confounded previous research through the use of the Raven's Standard Progressive Matrices Test combined with an extremely "user friendly" computer, the Apple Macintosh, programmed specifically to provide graphic feedback and directions. In addition, examiners provided supervised mastery of the testing format prior to the presentation of scorable items. The test results of this study indicate that these

modifications overcame many of the aforementioned limitations of the above research.

Two final APAS studies will be considered in detail, due to their recent publication date, relatively representative nature of APAS hardware and the inherent limitations of such equipment. In addition, these studies automated the RSPM, the same instrument used in this study.

Calvert and Waterfall (1982) evaluated the effectiveness of an APAS administration of the RSPM with that of the manual version. They used a test-retest format with an $N = 92$. Their equipment was quite representative of other APAS equipment. They presented the 60 RSPM items on 35mm photographic slides on a back-projected 12 inch square, vertically oriented screen. A Kodak carousel projector presented the items in ascending order of difficulty. An eight button keyboard oriented in the same manner as the RSPM items was in front of the screen. The logic circuits used to control presentation and scoring of the subjects' responses were contained beneath the keyboard. A control panel was placed out of sight of the subject.

Those subjects given the manual version of the test were instructed as detailed in the test manual (Raven, 1960). Those being administered the automated version were asked to read a page of typed instructions attached to the machine. Retesting was performed approximately four to eight weeks after the initial test, wherein the same procedure was used.

The automated version used a digital display to present the score of each of the five sets of matrices. In addition, the total score and the time taken to complete the total test were recorded. The manual versions were scored according to the instructions contained within Raven (1960).

During the first test administration, fifty subjects were given the automated version while thirty-three were given the manual version of the test. The automated group mean was 42.4 while the manual group mean was 44.7. This resulted in a non-significant difference at the $p < .10$ level. The average time taken for administration of the automated version was 20.4 minutes and the manual version required 23.3 minutes. These differences were found to be non-significant.

The retest phase of the research suffered considerable subject attrition. Thirty-one subjects (37%) of the initial sample were available for the retest. Four groups were compared in the test-retest design: automated-automated (N=12); automated-manual (N=6); manual-automated (N=10); and manual-manual (N=3). No significant difference was found between any groups. Thus Calvert & Waterfall (1982) asserted that this automated version of the test was comparable with the manual version. They stated, as previously mentioned, that shy individuals preferred the automated version and others were simply intrigued by the machine. Calvert & Waterfall asserted:

"the aim of reducing the amount of clinician involvement in the basically clerical task of test administration has been achieved. The system's only requirements of the operator are that he be capable of switching it on at the mains plug and of seating the subject before it" (p. 309).

The rather flawed nature of this research, whose very small retest sample sizes almost assured the lack of significance suggests that this conclusion was somewhat overstated.

Watts, Baddeley and Williams (1982) performed a similar analysis of the RSPM to that performed by Calvert & Waterfall (1982). Their equipment was very similar. However, they utilized a tailored testing format for the presentation of the RSPM and an additional Vocabulary test (the Mill Hill Vocabulary Test). The tailored testing approach involved the selective abbreviation of the testing procedure, not by dropping subtests as is commonly done in abbreviated manual test presentations, but by systematically skipping items. Thus in this study

..."Items [were] tested by testing first Question 1, if this was correct moving on to Question 3, then to Question 6, followed by 10, 11 and 12. If a subject made an error, then the prior item in the series was presented. If this were correct, then the upward-stepping procedure was resumed. If it were wrong, then any intervening questions were given before performance on this set was terminated and the next set started with Question 1. For any given set, the subject was given a score determined by the highest question correct, minus any wrong answers en route."(p.334)

The present study did not employ a tailored testing format due to Raven's proscription of changing the order of item presentation. In

the RSPM manual Raven (1978) states:

"Research with the ordering of items over a number of years has demonstrated that reordering items which appear to be out of sequence does not necessarily achieve the desired result. It is essential to note that the sequence of items has a teaching function so that later items are solved successfully due to the experience provided by earlier items in any one set. To transpose a later item to an earlier position because it is solved with relative ease [as in tailored testing] can result in making the item harder to solve and at the same time fail to provide the sequence of learning which is the essence of the progression as originally conceived"(p. SPM9.)

Watts et al. (1982) did not address their failure to comply with Raven's standard administration procedure in their tailored testing manner of presenting the RSPM, although the quotation above would suggest that increased item difficulty in the automated version may have ensued due to the use of this item presentation format.

In their comparison of the automated, tailored version of the RSPM vs. the manual presentation, Watts et al. (1982) found no significant difference between the four groups assessed in the test-retest format: pencil and paper twice PP/PP (N=20); Automated Test twice AT/AT (N=20); PP/AT (N=20); and

AT/PP (N=20). The mean scores for the type of administration, however, did yield a significant difference in favor of the manual means of presentation. The PP mean was 45.73 (N=40), while the AT (N=40) mean was 40.35 for the first administration. The same approximate five point difference was found on the PP mean for the retest administration (PP mean = 46.38 vs. AT mean = 41.15). Watts et al. (1982) asserted that "It is clear that the advantage enjoyed by the standard presentation procedure holds for both easy and difficult items and is characteristic of all five sets [of matrices]" (p.335).

Thus, a consistent finding is that automation of manual tests, when scored using manual norms, prove to be more difficult than the original manual test. (Elwood, 1969, 1972a, 1972b; Overton & Scott, 1972; Knights, Richardson & McNarry 1973, Calvert and Waterfall, 1982; and Watts et al., 1982). Watts et al. (1982) accounts for this finding within their research with two related factors. They stated that the principal factor was:

"Unfortunately, when we designed our program we were unaware of the guidelines suggested by Raven for the

automation of his test. These include a suggestion that: "the person tested should not be moved onto the next item as soon as they have considered their choice, but it should be possible for the testee to delay for as long as they wish, correcting their choice if they wish to do so and finally moving on to the next item by making a separate response to show that they are ready to move on. During the delay period there should be a clear indication to the testee showing which item has been selected."(p. 339)

A related factor, stemming from Watts et al.'s (1982) lack of knowledge regarding Raven's recommendations for automation of his test, is the negative correlation found between speed of response and correct score. This may have resulted from subjects inability to consider the accuracy of their response after making it. Thus, an incorrect impulsive response, which under Raven's two component response recommendation would have been considered by subjects prior to their required second response, may have been changed to a correct response. Lack of this two component response (which exists within the manual format and was included in this dissertation's

research) may have produced the increased difficulty detected by Watts et al.

An additional factor, not considered by Watts et al. (1982), may have been the lack of indication or feedback demonstrating which item the subjects had selected. The failure of the APAS test equipment to provide feedback regarding which item the had subjects selected may have caused the automated version of the test to be more difficult, leading to the aforementioned consistent five point disparity between the automated and manual versions, regardless of order of test administration. This assertion is supported by the above mentioned findings of Brierley (1971), who reported that some of his subjects would have preferred [perhaps required] more feedback. In addition, Raven's recommendations suggest that subjects should have their tentative responses (the initial component of the two component response) clearly indicated to them.

Knights et al. (1973) and Overton & Scott (1972), both of whom used the Raven's CPM with retarded subjects, reported the need for more explicit directions. Knights et al. indicated that they would in future work revise the initial item to make it more comprehensible

to the retarded children with whom they were working. Thus, it would appear that these reported difficulties with test directions may have stemmed from the authors' apparent lack of awareness of Raven's recommendations for a two component response per item (initial selection and a second indication that the subject was ready to move on to next item) and feedback clearly indicating which item the subject had selected. Neither Knight et al.(1973) or Overton & Scott's (1972) APAS equipment appeared to provide these recommended response interactions.

In considering APAS research one must recognize that the nature of the equipment severely restricts the degree of interaction that the subject can have with the machine. All of the above research, with the exception of Elwood's (1969, 1972a, 1972b) work, involved test items presented on slides by projectors and tape recorders. Some form of keyboard or other means of eliciting an operant response from subjects was provided. After the subjects responded, the next item was presented. The fact that the test-retest comparisons in all cases were not significant when compared directly with manual versions of the tests, reinforces many

of the authors' assertions regarding the rote and mechanical nature of many of popular and currently widely used manual tests.

The automated versions of the various tests examined were only found to be more difficult than the manual versions when sophisticated statistical analysis of the order effect was performed by certain authors. (Watts et al., 1982; Knights et al., 1973) The increased difficulty of automated tests was also demonstrated by significantly fewer subjects reaching a basal level in Overton & Scott's (1972) research with retarded adults. In addition, Elwood (1972a) found that the magnitude of practice effect was significant in diminishing scores of those subjects given the automated version of the WAIS first, again apparently due to its greater difficulty relative to the manual version.

The relative inability of the APAS versions of automated testing to provide high levels of feedback and interaction with subjects surprisingly has not diminished their capability to compare favorably with manual tests. Consideration of the developments in Computer Assisted Assessment (CAA) both previously and in the current study will demonstrate that the advent of automated

assessment devices for the screening of mental processes which produces results completely comparable to manual versions of the same tests is rapidly be approaching.

2. COMPUTER ASSISTED ASSESSMENT (CAA):

This review of CAA will focus on microcomputer mediated approaches, since the proposed research will be performed on such a device.

Elithorn and colleagues (Elithorn, Cooper & Lennox, 1979; Elithorn, Powell, Telford & Cooper, 1980) have developed an automated presentation of the Perceptual Maze Test. The Perceptual Maze Test was developed by Elithorn in 1963, (Elithorn, Kerr & Jones) as a variation of the Porteus Maze Test. It was initially designed for use on minicomputers. However it has now been modified for use on microcomputers. Elithorn et al's research was quite concerned with the extremely accurate measurement of the response time of subjects involved in making a series of binary (right or left) decisions to solve the Perceptual Maze Test. Their research will be considered in detail under the forthcoming section pertaining to response time. The essential aspect of this research is that it measured human performance and decision-making/problem-solving skills by providing a task whose measurement properties were not available prior to the advent of CAA.

Acker (1980a, 1980b, 1983) has developed a microcomputer based neuropsychological assessment battery and an individual test similar to the Wisconsin Card Sorting Test (Grant & Berg, 1948) which he has employed for evaluating the degree of non-verbal deficits within chronic alcoholics. In his 1983 study, Acker addressed many of the issues pertinent to current research.

Acker (1983) states that he made no attempt to adapt already established manual tests to the computer, "since adapting established tests would push a small computer to its limits" (p.363). He also indicated that computer based tests had inherent constraints and strengths. Acker (1983) indicated that some of the advantages of CAA:

"... are that stimulus and response timing is precise and effortless and that geometric designs and sequencing of trials may be determined according to mathematical laws"(p.363).

Acker (1983) suggested that the advent of computerized psychological tests had to await the introduction of inexpensive microcomputers which were simple to operate and had large dealer networks to sell and maintain the machines. He asserted that the conventional computer keyboard is unsuitable for subjects' responses.

"This is because a computer keyboard would confuse patients as they searched for the appropriate key to press with each response"(p.362). Thus he developed an optional nine key response board, whose keys were alternately covered or exposed as appropriate for specific tests. Obviously the use of no keyboard, as in this dissertation's research, avoided confusing subjects to an even greater degree than the simplified keyboard described by Acker (1983).

Acker's earlier research (1980a, 1980b) used the Maudsley Automated Psychological Screening Tests (MAPS). He modified these in his 1983 work and developed the Bexley-Maudsley Automated Psychological Screening Tests (BMAPS). This more recently developed battery, which was derived from the MAPS, will be discussed.

The BMAPS consists of five subtests. The first requires the subjects to make right-left discriminations of a mannikin figure. Subjects respond by indicating in which hand the figure is holding an object. The computer rotates the figure showing it facing the subject or with its back toward the subject and also rotated it in an inverted or conventional standing position. A second subtest requires the subject to perform a symbol-digit task similar to Wechsler's

digit-symbol subtest. A third visual perception analysis task requires the subject to identify the odd design from an array of three. The fourth test provides a measure of verbal memory by initially presenting a word to the subject, requiring him to make a semantic judgement (state whether the word is a strong or weak word) in an effort to assist the subject to memorize it. After completion of the presentation of all of the words to be memorized, the subject is presented with an array of three words and has to indicate which was the memorized word. A final visual-spatial memory test involves the presentation of single designs which the subject is asked to memorize. Then an array of three similar figures are presented and the subject must pick out the memorized figure from the two distractor items.

Acker (1983) has developed a test called the Bexley-Maudsley Category Sorting Test (BMCST), which is similar to the Wisconsin Card sorting test and requires subjects to determine rules for sorting the designs displayed on the computer screen. Four different categories of designs are shown at the top of the computer screen and the subjects determine to which of the four design categories an item

at the bottom of the screen belongs. Feedback is given for **CORRECT** or **INCORRECT** responses. Thus the capability of CAA tests to provide feedback in response to subjects behavior is demonstrated by this test.

Acker (1983) stated that there are two obvious advantages to this computerized version of the test versus the manual administration of the Wisconsin Card Sorting Tests or its variations (Nelson, 1976)

The first is that rule-learning tests like the Wisconsin Card Sorting Test and the BMCST are very difficult to administer in the traditional way. This is so because patients unable to solve the test intellectually (a) often become irritated by having their poor performance observed by another person (the tester) and abandon the test or (b) learn to modify performance by the sighs and lifted eyebrows of the tester rather than 'correct/incorrect' feedback. A computer is not so threatening to patients, and it gives no unintended emotional feedback.

The second advantage is that the analysis of patient errors in the BMCST, although quite complex and time-consuming to do by hand, is trivial for the computer. For example, each incorrect response a patient makes is scored according to the following criteria:

1. Was this rule used in the previous sort?
2. Has this rule been used incorrectly in the previous five sorts?
3. If the last sort was correct, has there been a rule change?
4. Has the patient used an irrational rule in this sort?
(p.368)

Acker's (1983) example provides a clear example of the enhanced psychometric sophistication afforded by the CAA format. The computer is actually interacting with the testee and providing feedback to him. Failure to provide these two factors were two of the major deficiencies cited in the APA5 research reviewed above.

Beaumont (1980) is reported by Thompson & Wilson (1982) to be producing a number of microcomputer-presented and scored versions of standard psychological tests. These include: the Wisconsin Card Sorting Test, the Synonym section of the Mill Hill Vocabulary scale, the Eysenck Personality Inventory, the Money Road Map Test, Raven's Matrices, the Digit Span Test and various language usage and spelling tests. He is reported to be standardizing the automated forms of these tests in an apparent effort to overcome the possible discrepancies that have been found by other researchers

employing norms from manual tests on newly developed automated versions.

Only one publication to date has been found which has applied CAA techniques to the Raven Progressive Matrices. Rock and Nolen (1982) investigated the feasibility and potential benefits of computer assisted assessment with children. They automated the Raven CPM on an Apple II Plus. In addition, they employed two disk drives, a thermal printer, Clock/Calendar Card and a Sony 17-In video monitor. The 36 figures of the CPM were programmed into the computer using Applesoft Basic and an Apple graphics tablet. A linking program presented the 36 items and maintained timing, test scoring, and reporting subroutines.

The subjects responded to the CAA presentation of the CPM by selecting "... each item by a color-coded key response matched to a color box placed above each response option displayed on the monitor" (Rock & Nolen, 1982, p. 41). The authors' description of the apparatus indicates no utilization of a two component response, or any awareness of Raven's recommendation for the use of such a response format.

The subjects were 4 girls and 11 boys who had been referred to the University of Washington Psychoeducational Center for psychological and educational evaluations. Their mean age was 10.33 (SD = 2.44) and ranged from 7 to 14. They were also administered a WISC-R as part of the experiment and demonstrated a Full Scale IQ range from 65 to 126 (mean IQ = 105.21, SD = 15.27).

No counter-balanced design was used in Rock & Nolen's reported experiment. The subjects were administered the automated CPM in addition to the WISC-R. The resultant scores were then correlated and the CPM vs. WISC-R intercorrelation was found to be $r = .59$ ($p < .05$). Rock & Nolan (1982) reported that:

"this correlation was similar in sign but smaller in magnitude than the coefficient obtained between the standard Raven and the Terman-Merrill scale, Form L ($r = .66$) reported by Raven et al. (1977) for a 9 yr-old British sample ($N = 209$)" (Rock & Nolen, 1982, p.42).

The relationship of the Terman-Merrill scale, Form L with that of the WISC-R was not discussed in the research report.

The authors also correlated the CAA based scores from each subscale (A, Ab, and B) of the Raven from the previously described sample (N = 15) with those of the normative sample reported by Raven (1977). They found the intercorrelations to be .53, .48, and .72 respectively. The authors report that "the magnitudes of the three intercorrelations were significantly different from zero ($p < .05$) and similar to, although smaller, than those reported for the normative sample" (p.42).

Rock & Nolen (1982) also compared the relative contribution of each of the aforementioned CPM subscales from the CAA with those of the normative group reported by Raven (1977). They found no significant difference between the contribution of the CAA versus the normative study group subtest scores toward the total score on the CPM in subtests Ab and B. They did find the groups differed significantly on subtest A, in the relative ranking of the this subtest's contribution to the total score.

Thus, Rock & Nolen's (1982) study demonstrated the same directional relationship on Subtest A as that of other automated

testing research using the APA5 format: i.e. the automated version scored with the manual norms was more difficult than the original test. (Elwood, 1969, 1972a, 1972b; Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982) The mean score for the experimental (CAA) version of the CPM was 11.9, while the normative version (Raven, 1977) was 18.32. The Mann-Whitney $U = 58.5$, $p < .05$) thus showed a significant difference. However, assumptions regarding the directional relationship based on Rock & Nolen's research is dubious. The other two subtest scores did not demonstrate the directional relationship which was consistently (although not always to a significant degree) found in other studies. The automated version of the CPM subtests Ab and B yielded scores which proved to be less difficult than the normative study group's mean scores.

"The mean ranks for subtest Ab were 16.1 (experimental) and 13.82 (normative) (Mann-Whitney $U = 88.5$, $p > .05$). Mean ranks for subscale B were experimental 15.2 and normative 14.67 (Mann-Whitney $U = 102.0$, $p > .05$)" (p.42).

Thus, unlike the other aforementioned studies, Rock & Nolen (1982) did not demonstrate the consistent, directionally based finding that suggested (generally not to a significant degree) that the APAS presentation of the various tests' items increased the degree of difficulty of the task, thereby lowering the mean scores of the automated version in comparison with those of the manually administered test.

Positing possible explanations for the disparate findings of Rock & Nolen's (1983) research relative to the other research findings must be strongly qualified due to the limitations of Rock & Nolen's (1982) study. Their study failed to use a counter-balanced, test-retest design as did the others. (Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982) The sample size of their study was quite small ($N = 15$), thereby reducing the possibility for the discovery of significant relationships. The results of their study employed the questionable practice of comparing their experimental results with those of Raven's (1977) normative study group. Further, in making this questionable comparison, Rock & Nolen (1982) compared their group,

which had a wide age range (7 to 14 years, Mean = 10.33, SD = 2.44) to the normative group using the latter groups' "...9-yr.-old British Sample (N = 209)"(p.42). The reasoning behind the choice of this normative age was not discussed in the article.

Rock & Nolen's (1982) study appears to be of somewhat limited generalizability relative to the APAS studies' (Elwood, 1969, 1972a, 1972b; Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982) due to methodological factors in addition to statistical concerns. They failed to employ Raven's recommendations cited in Watts et al. (1982) for a two component response format or the recommended feedback wherein the subject's initial (and tentative) responses would be displayed.

Nevertheless, beyond these qualifications, it may be suggested that the enhanced capabilities of the CAA format of automated testing could have been a factor in the findings which suggested that automated testing did not induce greater difficulty than the manual version. Clearly as Rock & Nolen (1982) assert: "these findings suggest that the potential utility of computerized testing

applied to psychological and educational assessment warrants continuing research investigation" (p.42).

C. RESPONSE TIME RESEARCH

The current research accurately measured response time as "it is generally accepted that measures of response latency are an important index of competence on a task" (Volan, 1982, p.303). Furneaux (1960) and Birren (1974) have also reported on the usefulness of response time in assessing psychometric parameters. Weiss (1979) contains numerous articles which emphasize the value of response time as well as response appropriateness. Response appropriateness will be discussed in the forthcoming section on 'motoric efficiency'.

Hunt (1978; 1980) and Weinman (1982) have examined the relationship of response time with other psychometric parameters in an effort to gain understanding of the nature of individual differences in cognition. Hunt (1980) has observed moderate correlations between verbal ability and several cognitive processes including speed of letter identification and speed of scanning for information in short-term memory.

It is well known that many psychometric tests are timed and often incorporate bonus points for responding quickly. Thus, speed of performance is a frequent determinant of individual differences. Eysenck (1967) has argued that item response time (the time for completion of each of the items within a test) is a meaningful measure of individual test performance. He asserted that if these item response times were plotted against item complexity, then individual performance slopes could be derived. In addition, he has suggested that for subjects of varying ability, the slopes would be parallel but with lower intercept values for the more capable subjects. Weinman, Elithorn & Farag (1981) have performed this type of analysis with the Perceptual Maze Test (PMT) and noted that it was possible to relate the performance slopes to chronological age in a sample of children between the ages of 8 to 17 years. Weinman (1982) asserted that his earlier (1981) finding confirms that more capable learners are faster but it

...does not begin to demonstrate how or why they are faster

In order to do this it is necessary to look in detail at the overall response time on a test item and to fragment it into

components which reflect different cognitive processes or different stages of processing" (p. 324).

Both Ellthorn, et al. (1980, 1982) and Weinman (1982) have researched this concept of "behavioural fragmentation" (Ellthorn, 1982, p.259) with automated versions of the PMT. This test was originally developed by Ellthorn, Jones, & Kerr (1963) as a paper and pencil test designed to measure a subject's ability to make binary decisions (whether to go right or left) while at the same time planning the next movements necessary to obtain the best solution, which involves intersecting the greatest number of dots. (Please refer to diagram #1.)

Ellthorn et al. (1982) stated:

"thus from a subject's performance on the automated version of the Perceptual Maze Test, it is possible to extract the following indices:

1. **Search Time:** the time until the first motor response.
2. **Track Time:** The time from the first motor response until the completion of the task.
3. **Check Time:** The time between a subject completing his tracking and his signifying that he is satisfied with his solution.

4. **Non-fatal Errors:** The Number of Incorrect movements initially chosen and then changed to a correct answer.
5. **Motor Index:** The average of fastest 10% of key response.
6. **Refresh Index:** The number of pauses >1 sec during the tracking phase.
7. **Laterality Index:** The percentage of right [direction] preferences.
8. **Processing Speed:** The number of vertices processed per second."
(Elithorn, et al, 1982, p.259)

This very detailed analysis of subjects' responses has been found by both Elithorn et al.(1979; 1982) and Weinman (1982) to be revealing of very subtle changes in subjects' physiological and psychological states. Elithorn et al. (1982) reported noticeable changes in one subject's response profile during modifications to their psychotropic medications. Weinman (1982) asserted that the PMT is sensitive to personality factors, "In particular, it was found that more extroverted subjects spend proportionally shorter time on the initial search phase [of the PMT]" (p. 325).

The CAA instrument developed for this study utilized all of the behavioral components assessed by Elithorn, et al. (1982) with the

exception of #5-Motor Index, #7-Laterality Index, and #8-Processing Speed. (Elithorn, et al, 1982, p.259)

Diagram #1

(Elithorn, 1982, p. 255)

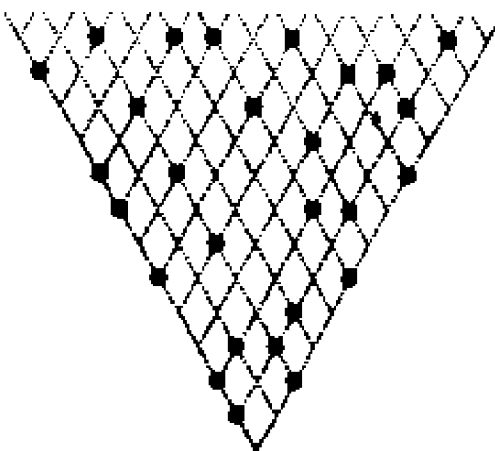


FIG. 3. The subject's task in the Perceptual Maze Test is to find a pathway along the background lattice which passes through the greatest number of target dots. He must keep to the lattice or tracks and must not cut across from one path to another. At each intersection the path must continue forward, i.e. the subject may fork right or left but must not "double back". In general, dependent on the arrangement of the target dots, there is more than one "best" pathway and the subject is said to have succeeded if he finds any one of these. There are two main conditions under which the Perceptual Maze Test is presented. A subject is either told the maximum number of dots which can be obtained, or this information is withheld and he is then left to decide whether he has found a "best" solution. Conventionally, these two methods of presentation are called the "with information condition" and the "without information condition".

These behavioral indices were not relevant to the current research due to the different type of task assessed by the RSPM. The PMT is a test which involves a subject making binary decisions of whether to move to the right or the left, in an effort to create a path that passes through the greatest number of dots. Thus, the PMT requires a Laterality Index to determine whether subjects have a response set favoring right or left. This study's newly developed CAA device allowed determination of the presence of a response set through the recordation and storage of all of the CAA subject's fatal and non-fatal errors for each item of the RSPM.

The Motor Index of the PMT proved inappropriate for inclusion in this dissertation's research due to the non-use of a keyboard. This dissertation's research contained a measure of 'motoric efficiency' which will be discussed during the next section of this research review. This linear measure of the efficiency the subject displayed in moving the mouse-mediated cursor from its initial starting point to an answer and back to the assigned area designated for "locking in" the subject's choice and moving on to the next item provided a measure similar to Elithorn et al.'s (1982) Motor Index.

The Processing Speed measure of the PMT is not appropriate to the RSPM as there were no real vertices. A measure similar to this was available as a function of the number of pixels (not visible to subjects, unlike vertices of PMT) transversed by the cursor per unit time. The three response times and two measures of motoric efficiency were measured in this study, although their interaction in a speed parameter does not appear to be meaningful to the task assessed by the automated version of the RSPM. Therefore no direct speed computation of this variable was made in this study.

Teizrow (1983) suggested that the use of response time and the related phenomenon of movement time were important measures of a child's neurological level of functioning. She asserted that these measures, coupled with the use of a measure of vocabulary acquisition, may meet the needs of a child neuropsychological assessment. She posed this recommendation as an alternative to scaling adult indices of neurological functioning down to children.

Very limited work in CAA or automated testing has been done with children as subjects. As previously mentioned, Knights, Richardson & McNarry (1973) assessed retarded children's

Intellectual potential using automated (APAS) versions of Raven's CPM and the PPVT. The aforementioned Weinman, Ellithorn & Farag (1981) and Telzrow (1983) research explored the feasibility of using automated testing techniques to assess the response time of children. Rock & Nolen (1982) to date appear to be the only researchers to have employed a standardized, manual test of intellectual development in a CAA form of presentation.

Thus, there appears to be a viable relationship between directly measurable cognitive factors such as response time and other psychometric parameters. Unfortunately very little work has been done in this area, particularly with children. The inclusion of a three component response time measure (scanning prior to initial movement of the mouse, time from initial movement prior to initial item selection, and time until 'locking in' the response in order to move on to the next item) as part of a standardized computer assisted intelligence test, offers an excellent opportunity to increase our limited understanding of the actual behavioral determinants underlying the psychometric concept of intelligence.

D. PERTINENT RESEARCH RELATED TO THE CURRENT STUDY'S MEASURE OF 'MOTORIC EFFICIENCY'

Watts et al. (1982) has cited two factors which limit the clinical usefulness of automated testing.

"One of these is the possibility that patients with mental impairments may not be able to interact with a machine appropriately. A second - and possibly greater - difficulty is that a major cue to mental function often lies in the quality of the response made by a patient to a question, or the manner in which he sets about a task, rather than in the content of the response itself"(p.331).

The first objection has been discussed in the section concerning APAS research. There the observation was made that in all of the reviewed APAS studies (Elwood, 1969, 1972a, 1972b; Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982) the findings suggested that the automation of the pre-existing tests, which had been designed and normed for manual administration, proved to be more difficult in the automated format. The added difficulty the automation of the WAIS produced, as reported in Elwood's various studies, (1969, 1972a, 1972b) appeared readily attributable to the

increased difficulty introduced when the manual (performance) tasks were automated. Limited understanding of the test directions was attributed to be a major difficulty in Knights et al. (1973); and Overton & Scott's (1972) research with retarded children and adults respectively.

In addition, it is possible that the increased difficulty of the automated versions of the APAS tests, in addition to the confounded findings of Rock & Nolen (1982), stemmed from the failure of the equipment to interact with the subjects. As one reads the descriptions of the apparatus of the various studies, one recognizes that the machines are simply presenting items and eliciting operant responses. No true interaction occurred between the subjects and the machines. Raven's recommendation that subjects be given feedback to indicate their initial responses and allow the opportunity to change their responses has apparently not been incorporated in automated test equipment prior to this dissertation's research. Prior studies would suggest that a CAA device which indicates graphically the machine's reaction to the subject's movements (such as with the "mouse") will enhance the subject's sense of interaction with the

machine. The subject's sense of interaction should be further enhanced if the device produces clearly depicted feedback of the subject's tentative selections and allows him to make changes to these choices as well. Consequently, the CAA device designed for this dissertation incorporated these added assessment stimuli in an effort to diminish the unsuitable aspects of CAA devices previously revealed in the literature.

This study initially proposed teaching the subjects to interact with the computer by using a series of non-scored items presented in a stimulus fading format. The initial impression was that such an error-free learning format would provide the subject with an opportunity to master the manner by which the CAA device required them to respond. Field testing of the prototypic CAA device which used only the actual RSPM items indicated that children readily grasped the CAA testing format when the manual version of the RSPM directions were used with practice allowed on the initial two items of the test (A_1 and A_2). Therefore, this researcher determined that the possibility of confounding results through "teaching the test" would serve little purpose and so the CAA device's introductory

directions were designed to closely parallel those provided to the subjects of the manual RSPM.

The use of novel, but psychometrically indicated elements of automated testing designed to increase the subject's interaction with the machine, should diminish the degree of difficulty involved in using the automated version of the RSPM. These factors should address in large measure Watts et al's. (1982) concern regarding a testee's inability to interact with a given research's apparatus.

Watts et al's. (1982) second concern, regarding the subjective, qualitative elements of psychometric testing will also be addressed, to some degree by this study's measure of 'motoric efficiency'. As discussed in the Introduction, 'motoric efficiency' is a two component, linear measure of a subject's movement of the pointer or cursor, which is controlled by the Apple Macintosh's 'mouse'. Thus a highly accurate quantitative measure of the child's movement of the pointer, both before his initial choice and from this time until he 'locks in' his response and moves on to the next item, will be provided. In this way, the degree of purposefulness of the subject's movements will be assessed during each item's selection,

as well as during the time in which the youngster has to finalize his decision. The potential exists to determine different problem solving styles when the interrelated, two-phase 'motoric efficiency' measures and three phase response time measures are analyzed in relation to the subjects' item responses (correct/incorrect). A quantitative measure of the child's tolerance to personally frustrating items may also subsequently be derived.

Watts et al's. (1982) concern that automated testing would diminish the psychologist's awareness of "the manner in which he [the subject] sets about a task" (p.331) would appear to be largely overcome, if not superseded by the inclusion of a quantitative measure such as this study's 'motoric efficiency' parameter.

In current, manually administered tests, the interpretation of a testee's manner of approaching the task has the potential to be strongly affected by the administrator of the test. For example, what a highly compulsive evaluator may report as a "dillatory child", to another evaluator may be "a diligent worker". There is generally no means to determine from test protocols the accuracy of these subjective determinations. Thus, future researcher's refinement of

this prototypic and relatively basic quantitative measure of a subject's response style may well increase the objectivity by which psychologists examine and report such subjective information as the manner in which the testee interacts with test materials.

Weiss (1979), as previously mentioned, contains numerous articles which call for the inclusion of response time and response appropriateness into CAA research. The inclusion of 'motoric efficiency' in the current study provided an indication of the appropriateness of the subject's responses. For example, should a subject move his cursor across the screen in a meandering manner, the 'motoric efficiency' measure will provide a precise assessment of his inappropriate responses. Since the 'motoric efficiency' measure is two-phased for each response, the examiner may obtain a detailed, quantitative measure of initiation (When did the subject begin non-purposeful movement - eg. during personally difficult items only?), frequency (On how many items did the subject exhibit this style?), and extent (To what extent did the subject wander about the screen non-purposefully, relative to other subjects?).

Thus numerous researchers have expressed concern for the loss of subjectively derived information about the nature of a subject's performance on a task. The current research considered here regrettably cannot produce subjective inferences about a subject's behavior. However it can introduce a measure which attempted to increase the degree of quantitative analysis of subjects' problem solving approaches, tolerance to frustration, purposefulness in responding and an indication of their degree of certainty in responding. These psychometric factors have rarely been quantified in past research. Ellithorn and coworkers (1979, 1980, 1982); and Weinman (1982) have initiated work on what Ellithorn referred to as "behavioural fragmentation" (1982, p.259). They evaluated this form of assessment using the PMT. This test required the subject to make a series of binary decisions. Thus the actual behavior analyzed was quite simple. Nevertheless, the researchers found that the CAA format was sensitive to drug effects and in revealing personality factors. Thus, the current research significant findings of relationships between the interaction of response times, motoric

efficiency', and the measures obtained in the post hoc analysis of the WISC-R was foreshadowed by the previous research.

E. RESEARCH REGARDING THE POTENTIAL EFFECTIVENESS OF THE APPLE MACINTOSH COMPUTER FOR CAA STUDIES.

The current research employed an Apple Macintosh computer which had 512 Kilobytes of Internal random access memory (RAM). This study initially considered the use of two floppy disk drives, one internal and one external, each providing 400 Kilobytes of memory storage per disk. However, as preliminary testing of the CAA device occurred it was determined that 800 Kilobytes of storage did not allow for adequate ease of administration or rapid enough processing of the stored RSPM images. Consequently, an internally mounted, 10 Megabyte (Ten Million Characters of Storage) hard-disk drive was employed. The use of such a device had earlier been ruled out due to cost and portability factors. However a compact, relatively inexpensive hard-disk drive called a Hyperdrive™ became available during the development of this study's CAA instrumentation. These hardware/software considerations are extensively discussed in Chapter Four. (Please refer to Chapter Four - Development of the Automated Version of the Raven Standard Progressive Matrices).

The 32-bit processing format of the Macintosh allows for relatively compact storage of the RSPM items (approximately 10 Kilobytes per figure). The relatively (compared to most microcomputers) large 512 Kilobyte internal memory of the Macintosh computer allowed for a large amount of information to be loaded into memory, without having to refer to the storage disk. It had been hoped that the next RSPM item could be loaded into memory, while the subject was responding to the previous one. Loading the next item while the subject worked on the preceding item would have minimized any difference between time of administration of paper and pencil versions of the RSPM and the CAA format of the proposed research. Unfortunately the single Central Processing Unit (CPU) format of the Macintosh, precluded this activity since inaccuracy would have been introduced into the measures of response time and motoric efficiency which are continually obtained as the subject responds to items. (See Chapter Four) Consequently, the significant findings of this study regarding time of test may have stemmed from this factor.

The bit-mapping method of graphic presentation utilized by the Macintosh allowed subjects to respond to the automated RSPM through the use of a device commonly referred to as a "mouse". The subject responded simply by rolling the "mouse" device on the table near the computer screen. A pointer or cursor on the computer screen moved in a manner corresponding to the direction and extent of the "mouse's" movement. The graphic screen display of 512 x 342 pixels per inch provided very detailed, high resolution black, varying shades of gray, and white images. The bit-mapping format allowed the Macintosh to be programmed to count the linear pixel by pixel movement of the mouse yielding the previously discussed measure of 'motoric efficiency'. As the reader will note, the high degree of resolution provided very accurate measurement of the subject's pointer movement, in excess of 80 units per inch of movement.

The Macintosh also offers mainframe compatibility. Consequently, eventual highly detailed, item-by-item statistical analysis or more advanced future CAA development is possible given the Macintosh's 32-bit design which allows for direct, microcomputer to mainframe communication.

Beaumont (1982) in his extensive discussion regarding the hardware requirements for CAA asserted: "A display of, say, 280 x 192 [pixels] will allow production of quite complex stimuli as found for instance, in the Raven's Progressive Matrices tests" (p. 313). In addition he also stated: "the graphics capability of the system is most likely to be a critical feature of the hardware specifications" (p.312). Regarding the memory capability of machines involved in CAA, Beaumont indicated that "a machine offering between 32k and 48k of core storage is likely to be sufficient for a reasonable range of testing programs" (p.312). Obviously, the Macintosh's graphic and memory capabilities far exceeds these recent hardware recommendations.

Beaumont (1982) also discussed the use of "Mice" in his article. He stated: "undoubtedly, the single most important topic in considering hardware for interactive testing is the selection of **response media**" (emphasis added). He went on to discuss the limitations inherent in using keyboards and indicated that "mice" offer an alternative to keyboards, yet to his knowledge none had yet been employed in CAA research (p.314).

Elithorn et al. (1982) discussed a concern that "since composing and presenting such complex displays with a small microprocessor is time-consuming, the paper-and-pencil version would probably take considerably less time to complete" (p.251). This assertion appeared to be correct as regards the significantly longer time required for the CAA version of the RSPM versus the manual version examined in this study.

Thus, in conclusion, the 512 Kilobyte Macintosh met or exceeded the recommended attributes of numerous researcher hardware requirements for CAA. The advertising slogan used to market the Macintosh, "If you can point, you can use a Macintosh", accentuate the ease of use which comes from the "mouse-based" format of user interface.

E. SUMMARY OF RESEARCH REVIEW

This research review has demonstrated the efficacy of automated (both APAS and CAA) testing formats. Generally, those studies which have compared manual versus automated administrations of the same tests have found no significant differences between the two types of test administration. Although no significant difference was generally shown in the majority of studies reviewed, a consistent finding (with the exception of the mixed results of Rock & Nolen, 1982) was that the automated testing groups' mean scores tended to be somewhat lower. This may have indicated an increased degree of difficulty stemming apparently from the automation of the test instrument relative to the manual version. The Rock & Nolen (1982) study failed to clearly demonstrate this finding. Numerous deficiencies of this study were discussed, the possibility exists that the confounded findings of this study stem from the study's use of CAA rather than the APAS format of test presentation. The other previous studies discussed had all used the APAS format. This assessment approach may have introduced greater

difficulty than either the manual or CAA versions of the tests which they had automated.

The findings of the automated testing research clearly indicated that greater diligence should be placed on adhering to the administration procedures prescribed by the test designers. In the specific case of the Raven matrices, (both CPM and RSPM) researchers who have automated the instruments failed to attend to guidelines pertaining to the nature of the subject's responses. Specifically these are:

1. The need for a two component response,
2. The need for clear feedback of which item the subject had selected and
3. Maintaining the order of test item presentation.

The latter factor had been violated during the use of a tailored testing format which mixed the order of item presentation, thus violating Raven's proscription against such modifications.

This study performed the first known measure of the efficacy of an automated RSPM, in a counter-balanced design employing a CAA (as opposed to an APAS) format versus a manually administered

version of the test. In addition, the study also appears to be the first known research which closely adheres to the recommendations of Raven, the test developer, in providing a two phase response format with clear feedback of the subject's initial and tentative responses.

The inclusion of a three-phase response time measure into the this study was discussed. Previous research suggests that response times may be a very useful measure in assessing children's cognitive capabilities and personality factors. The widely used application of response time measures in intellectual evaluations was also discussed. This study's findings of significant relationships between response time and other psychometric parameters had a sound research basis. (Please refer to discussion in Chapter Five - Results)

The development of this study's measure of 'motoric efficiency' was discussed. This measure attempted to overcome a major concern of some psychometric theorists, i.e. that automation does not provide subjective information regarding the response style the subject employed in solving the various test items. The measure of 'motoric efficiency' provided a highly accurate, quantitative depiction of each testee's response style. The different response styles demonstrated

by subjects in this study was found to significantly relate with other psychometric parameters. The further development of this two-phased factor, coupled with the three-phased measure of response time, may provide examiners with greater information about a testee's test behavior than would be available to examiner's employing manual versions of the same tests.

The current study's analysis of small fragments of behavior may expand the work currently being done by previously discussed workers using the PMT. The great potential of further CAA research in precisely measuring the actual process of, and individual differences involved in, problem solving offers tremendous potential to psychologists interested in modifying human behavior as well as those concerned with assessing the distribution of psychometric parameters.

Finally the research review considered research regarding hardware requirements for CAA research. The researchers' recommendations all suggested that this study's use of the Apple Macintosh met or exceeded the requirements for graphic depiction, memory and manner of subject - machine interface. Consequently the

numerous significant relationship found using this device had a strong research basis.

CHAPTER III

METHODOLOGY

A. Population and Selection of the Sample

This study consisted of two different evaluations of Computer Assisted Assessment (CAA). The first phase of the study examined the reliability and concurrent validity of the automated Raven Standard Progressive Matrices (RSPM) with the manual version of the same test. A second phase of the study investigated the psychometric relationships between previously administered WISC-R tests and the automated RSPM. The population consisted of upper elementary students for both portions of the study.

The first phase of the study attempted to obtain a sample of sixty (60) randomly selected students, selected from the 4th, 5th, and 6th grades of two elementary schools. In attempting to obtain the projected N=60 for this group of subjects, this experimenter allowed for an attrition factor of 40%. The selection of the subjects for the first phase of the study was performed randomly. The number of students in each of the two selected school's fourth, fifth and sixth

grade levels was determined and a common denominator which would yield 14 students from each grade level at each school was determined. The experimenter then counted out every seventh student on the 4th, 5th and 6th grade rolls in one school or every eighth student on the roll in the other, larger school, and selected them for participation in the study. One selected student, who attended an Emotionally Disturbed Special Education class, was not included in the study due to safety factors stemming from the experimental requirement that subjects operate the automated CAA equipment in an unsupervised environment. Finally, 14 students from each of the Fourth, Fifth and Sixth Grade levels were selected from each school for inclusion in the first portion of this study.

This first group of subjects were given parent permission forms (Please refer to Appendix B) which explained the nature of the research and had a tear-off section for the parents to sign and return indicating their permission or refusal for their child to participate in the study. The return rate for this group of subjects was very good, slightly exceeding the projected return rate of 10 out of 14 requests per grade level per school.

Students were assigned to the three experimental (Manual - CAA; CAA - Manual; CAA - CAA) groups by randomly drawing cards with the students' names and sequentially placing them in one of three stacks. Each stack was then randomly assigned to one of the three test conditions. Subjects were then administered the first portion of the counter-balanced test-retest format of this experiment's design. Following a two-week interval, they were administered the second portion of their assigned testing group.

The second phase of the study involved students in the 4th, 5th and 6th grades who within the 1984-85 school year had been administered a WISC-R by a certified school psychologist. The names and respective schools of these youngsters were obtained by a questionnaire from this experimenter to each of the fifteen school psychologists working within the School Division in Southeastern Virginia in which this study took place. This questionnaire was issued following approval from the Director of Pupil Personnel Services this school division. It requested psychologists to determine their highest referring elementary school and list the children within this school who had been administered the WISC-R during the 1984-85

school year. In addition, they were asked to also list the reason for referral and the date of testing of each child. A permission letter was sent under the auspices of the Department of Research and Testing to each of the principals of the selected elementary schools indicated by the participating psychologists. The nature of the research was described and a tear-off portion of the form was provided to allow these principals to indicate their permission or refusal for their school and assigned school psychologist to participate in this research (Please refer to Appendix B).

Fourteen psychologists obtained permission to participate in the study. They provided a list of student's names who had been given the WISC-R within the preceding twelve months of this study.

Three students from each of the participating fourteen psychologists' lists were randomly selected. Choosing three subjects from each psychologists' pool of potential subjects built in an approximate 50% attrition level. In addition, randomly selecting the same number of subjects from each psychologists' pool of potential subjects attempted to equalize examiner effects. This equalization occurred in limiting each psychologists' contribution of subjects to

one elementary school (number of assigned elementary schools vary among psychologists) and to three potential subjects per psychologist. Thus assuming a 2 out of 3 participation rate a sample size of $N=30$ appeared likely for the post hoc comparison of the WISC-R with the automated RSPM.

The unanticipated high degree of transience among the selected pool of subjects (numerous potential subjects had moved from the school area in which they had previously been given the WISC-R) and a number of parents who refused permission for their child to participate in the study, severely diminished the number of subjects within this portion of the study. A sample size of $N=20$ was successfully administered the automated version of the RSPM from this group of 45 potential, randomly selected students. Ten of the initial fourteen psychologists administered the test to subjects. One psychologist administered the RSPM to one student and one other psychologist administered this test to three subjects. The remaining eight psychologists each administered two automated RSPMs to students within their highest referring elementary school.

B. Treatment and Data Gathering Procedures

1. Data Gathering

The subjects were administered the automated RSPM in a procedural and graphic format as close to the manually administered test as given according to the published standardized directions. Previous research has established that no significant difference exists between automated simulations or mainframe computer administrations of the RSPM and the scores of the manually given version. The subject's responses were recorded and scored in the manner prescribed by the RSPM test manual.

The familiarization process offered by the directions provided an opportunity for the subjects to learn to interact within the CAA testing environment prior to beginning the actual test. Items A_1 and A_2 (first two items on the RSPM) were used as non-scored, learning opportunities for the CAA version of the RSPM just as they are in the standardized paper and pencil version of the RSPM. Demonstrated mastery of these preliminary items under the supervision of the examiner ensured that the subject understood the nature of the CAA

task. The CAA device was constructed so that the examiners could selectively repeat items A_1 or both A_1 and A_2 to provide greater reinforcement of the skills and strategies involved in using the CAA version of the RSPM. When the examiners determined in their judgement that the child understood how to use the CAA testing format, the subject was allowed to continue with the test without supervision until its completion. At this point the examiner unplugged the Macintosh's keyboard and the subject started to work on the actual items of the RSPM. This phase of his work was unsupervised (overtly, atleast), since lack of the need for supervision is an essential purpose underlying the development of CAA.

For each item of the test, the subject was initially presented with a screen depiction of that item of the RSPM, accurately drawn on the computer screen. The textual prompt 'move your arrow to the correct answer and click' was shown at the top of the screen, encouraging the subject to make his choice. After the subject clicked the mouse on the answer of his choice, that response was highlighted by a black rectangle which appeared around the chosen item, and a new prompt appeared at the top of the screen. It stated: 'check or

change your answer and lock it in here ->'. This prompt's '->' pointed to an enclosed rectangular area in which the subject had already practiced "locking in" his responses. Some degree of eye-hand coordination was required for the second phase of the subject's response. The initial portion only required that the subject place the cursor relatively close to the answer of his choice, which was sensitized by the computer's assessment program to note his choice and highlight it. Locking in the response for each item by clicking the mouse button in the designated area evoked the computer to promptly refresh the screen with the next item of the RSPM. This process continued until the subject completed the entire evaluation.

The CAA version was also equipped with a means of terminating the test prior to its completion, should a subject refuse to continue. No subjects from either the CAA or the paper and pencil groups failed to complete all items of the RSPM.

The computer was programmed to maintain each subject's record of responses and all of the other previously discussed CAA psychometric parameters (eg. the measures of motoric efficiency and response time). Each subject's file also included their name, the

examiner's name, school, date of birth and the date of the test. The clock time of the test was automatically recorded from information provided by the internal clock within the Macintosh computer. This data was stored on the hard-disk drive in a file which was automatically created for each subject at the beginning of the test. These files were kept apart from the CAA program which mediated the administration of the RSPM. Complete confidentiality of the recorded scores and all psychometric parameters was maintained.

Two measures unique to computer-based assessment were taken. The child's item response times and cumulative response times were measured and recorded in milliseconds. A measure of 'motoric efficiency' was made through the linear measurement of the child's movement of the cursor (through mouse control) across the screen on both a per item and cumulative basis. These results were analyzed according to well accepted psychometric and psychophysical approaches reported in the literature.

2. Ethical Safeguards and Considerations:

The current study was reviewed and approved by the human subjects review committee of The College of William and Mary. In addition, the Research Department of participating school division reviewed the proposal prior to authorizing the use of school division students in the research project. These reviews, in addition to the thorough review by the Doctoral committee supervising this dissertation helped to ensure that the planned research was performed according to sound ethical principles.

Parental permission was obtained through the use of a letter reviewed by the dissertation committee and the Virginia Beach Research Department. Two different letter formats were used, one for each part of the study. The first letter was sent to parents of subjects who were selected to be administered the manual and automated versions of the RSPM during the counter-balanced, test-retest portion of the study. This letter informed parents of the opportunity for their children to engage in research and experience interacting with a computer for approximately one hour. In addition,

parents were informed that a measure of their intellectual potential would be obtained, but that, due to the experimental nature of the study the results would not be made available to anyone except the researcher. The second parent permission letter was sent to parents of subjects involved in second part of the study, the post hoc comparison of the automated RSPM with the WISC-R. This letter was similar to the first with the exception that permission was also obtained to refer to the subject's existing confidential records. Complete confidentiality and anonymity was assured and maintained.

This study's determination not to report CAA obtained test results to parents or educational professionals was based on the prototypic nature of the CAA instrument. Since a major goal of the proposed research was to determine the validity and reliability of the newly developed instrument, the accuracy of the test results must be considered questionable, until the conclusion of the experiment at the very least.

Beyond these guidelines, the issue of automated or CAA based evaluations poses several ethical problems, both currently and for workers in this area in the future. These issues center around the

eventual appropriate usage of such automated test equipment; i.e., who should oversee the administration of tests, what legal precautions must be taken to ensure informed consent, and who should have access to, and make interpretations of, test data.

It seems appropriate to assume that the ethical considerations for this CAA evaluation and for those in the future should be considered in the identical manner as any psychoeducational evaluation. Full parental permission, given after being fully informed about the nature of the evaluation should always be obtained. Complete confidentiality regarding the test results is required due to the sensitive nature of the information regarding subjects' intellectual potentials and other psychological data. As the research review has demonstrated, CAA offers the potential to generate even more accurate indications of intellectual and personologic factors than current manually administered tests. Due to these factors, interpreters of CAA derived test data may require an even greater degree of statistical and clinical training than is currently required of interpreters of present psychological test data. Thus, it seems clear that the minimal level of expertise required to interpret CAA

results should be certified school psychologists or other professionals with advanced training in appropriated areas of psychology.

One of the most promising aspects of CAA research and development is that the use of such technology will free highly trained professionals from the relatively clerical tasks of test administration with no significant loss of diagnostic information. Thus, the procedural task of assisting subjects to perform CAA should emphasize skills in which paraprofessionals may be trained, in an effort to enhance the productivity of psychologists', i.e. provide more time for them to directly interact with clients, rather than assess them. However, as previously mentioned, ethical guidelines of the profession should ensure that only professionals qualified to do so, interpret the data obtained by the paraprofessional and CAA based psychometric instruments.

C. Instrumentation

I. Description:

The RSPM was chosen for automation in this study due to a variety of factors. The most important factor was the feasibility of accurately generating the test item figures on a computer screen, in a format consistent with the published means of administering the manual version on the test. Previous research had demonstrated that this was possible, however the graphic quality of the computer presented figures was reported to be "unacceptable" (Raven, 1985).

Previous research on the RSPM indicated that it was "a useful measure of nonverbal reasoning ability" (p. 247). In addition it is reported to be a "culturally reduced test" (p. 248) and is very well adapted for children who have limited English skills (Sattler, 1982). Anastasi (1982) indicates that the RSPM is regarded as one of the best measures of Spearman's g factor (Spearman, 1927). Anastasi further indicates that the RSPM is a useful test for children who have severe language, auditory or physical handicaps.

Kaufman (1979) concurs with Anastasi's findings regarding the usefulness of the RSPM with physically handicapped or "motor-impaired" (p. 37) children. In addition he states that the RSPM is a capable test for assessing children's non-verbal intelligence. Kaufman suggests that the untimed nature of the RSPM is particularly beneficial for children whose scores are impeded on tests like the WISC-R performance section which employs overt timing of responses.

Thus the RSPM was selected due to its relative ease with which it could be automated, its rich research foundation and due to the demonstrated robust capabilities of the test instrument to assess varying populations of children in a relatively culture-fair manner.

The administration of the RSPM was mediated by an Apple Macintosh 512K computer with an internally mounted Hyperdrive™ hard disk drive. Storage of the test stimuli and program required approximately a total of 1,100 Kilobytes of storage. Each plate of the RSPM test items require 13 Kilobytes of memory.

The subjects responded to test items through use of the "mouse" which provided a means for the child to point to his choice

of the appropriate response. No keyboard was present during the automated testing process in an effort to diminish distracting effects and avoid introducing uncontrolled effects stemming from subjects' differential previous exposure to keyboards.

The Macintosh was chosen due to its excellent graphic depiction and graphic duplication capabilities (through existent software for drawing the RSPM figures or secondary source visual scanners designed to interface with the bit mapping format of the Macintosh). In addition the 32bit architecture lends itself well to mass storage of extensive amounts of graphic materials. The input device commonly called a mouse also allowed a reliable means for the subjects to respond, in a manner similar to the manually administered version and also yielded the additional psychometric parameters of response time and 'motoric efficiency'.

2. Reliability:

The reliability of the newly developed CAA version of the RSPM was determined during the first phase of the research. A test-retest comparison of the CAA - CAA test was performed with an intervening time of at least two weeks to diminish any practice effects which

might have ensued. Appropriate statistical analysis was performed to determine the degree of reliability found in the new CAA version of the RSPM. The CAA version's level of reliability was compared with that of the manual version of the RSPM and found quite comparable. (See Chapter Five - Results)

3. Validity:

The concurrent validity of the newly developed CAA version of the RSPM was compared with that of the manual version of the RSPM in a test-retest, counter-balanced comparison. The results were compared and found quite comparable.

The second portion of the research examined the post hoc concurrent validity of the CAA version of the RSPM with that of the WISC-R. This comparison examined the WISC-R's verbal, performance, full scale, subtest, factor analysis derived scores with the psychometric parameters of the newly developed CAA version of the RSPM. The IQ scores obtained by the automated test were obtained from the norms of the manual version of the RSPM. The published, English percentile norms were converted by interpolation to standard scores to allow for direct linear comparison with the

WISC-R IQ scores (See Chapter Five - Results). These comparisons determined the degree of concurrent validity between the two instruments. The correlation coefficient of the automated RSPM also compared with that of the widely researched manual RSPM, to determine the similarity between the two devices.

D.Design:

The first portion of the study was a counterbalanced design. It was designed to examine the effect of test-retest and practice effects of the three groups within this portion of the study. These groups are Group A CAA first - Manual second; Group B Manual first - CAA second; and Group C CAA - CAA. The test - retest of the manual version of the RSPM was not be evaluated in this study, as it has been found to be quite reliable by numerous studies.

The diagrammatic depiction of the first portion of the study.

	Time 1	Time 2
Group A	X ₁ O	X ₂ O
Group B	X ₂ O	X ₁ O
Group C	X ₁ O	X ₁ O
X ₁ = CAA Version		X ₂ = Manual Version

The second portion of the study was a post hoc or "ex post facto" experiment.= (Campbell & Stanley, 1963, p. 70). This portion of the study examined the capability of the CAA version to predict, in a post hoc

manner, the scores of the previously administered WISC-R subtest, verbal, performance, full scale and factor analyzed derived scores.

E. Specific Hypotheses:

1. There will be a non-significant difference between the computer assisted administration (CAA) of the RSPM and the manual administration of the the same test when compared in a counter-balanced, test-retest design.
2. The computer assisted administration (CAA) of the RSPM will demonstrate a non-significant difference between pre-test and post-test scores [using an appropriate mean comparison].
3. The computer assisted administration of the RSPM will demonstrate a statistically acceptable degree of reliability when measured in a test-retest format [using a reliability coefficient].
4. The computer assisted administration of the RSPM will not show any significant order effect (CAA – then - Manual versus Manual – then - CAA) in appropriate statistical comparisons.
5. Automated (computerized) administration of the Raven Standard Progressive Matrices (RSPM) will demonstrate a significant relationship in a post hoc manner with the outcome of formal psychoeducational evaluations using the WISC-R.

6. The automated RSPM will yield a degree of concurrence with the WISC-R scores comparable to previously established levels of concurrence reported in test literature for the manual administration of this instrument.
7. The results of the automated testing procedures will yield significant inter-relationships with the results of previously administered WISC-R subtest scores, or Verbal, Performance, and Full Scale scores when analyzed using the appropriate statistical procedures.
8. The results of the automated testing procedures will yield significant inter-relationships with the previously derived WISC-R subtest factors derived through factor analysis.
9. The unique psychometric parameters of subjects three-phased Item and total response times and two-phased measures of 'motoric efficiency' will provide statistically significant quantitative interrelationships when compared with other group and individual data obtained through automated and/or previously administered manually obtained psychometric data.

E. Statistical Analysis:

The evaluation of the obtained results relative to the previously stated hypotheses were made using a variety of statistical techniques (See Chapter Five - Results). An Alpha or level of significance of $p < .05$ was employed in making all the aforementioned comparisons. The results are reported in tabular form and their implications will be discussed in Chapter Five - Results).

G. Summary of Methodology:

The current study developed the programmed software necessary to present, administer and score the RSPM through the medium of an Apple Macintosh 512 Kilobyte microcomputer with 10 Megabyte Hyperdrive™ Hard disk drive. This research is the first known CAA (as opposed to APA5) study to automate the RSPM and the first study to employ the recommendations of Raven et. al. (1977, 1978, Watts, Baddeley, & Williams, 1982) for utilizing a two component response format and providing subjects with feedback of which answer they have selected during their tentative initial answer selection.

The first phase of the study was a counterbalanced comparison of the degree of concurrence between the manual vs. the newly developed CAA version of the RSPM. In addition, this portion of the study determined the reliability of the CAA version of the RSPM.

The second portion of the study evaluated subjects who during the 1984-85 school year had been administered the WISC-R as part of an evaluation to determine eligibility for special education. The subjects were selected at random from a list of students provided by the Director of the Pupul Personnel Department. A post hoc statistical comparison

between the previously administered WISC-R scores and the CAA version of the RSPM was performed to determine the extent of concurrence and predictive capability possessed by the automated RSPM relative to the WISC-R.

CHAPTER IV

DEVELOPMENT OF THE AUTOMATED VERSION OF THE RAVEN STANDARD PROGRESSIVE MATRICES

This research initially explored the use of an Apple II Plus computer in an effort to further Rock and Nolen's (1982) study and automate the Raven Standard Progressive Matrices (RSPM). Rock and Nolen's (1982) work automated the Raven Colored Progressive Matrices (CPM). They failed to employ the two phase response format recommended by Raven which was cited in Watts et al. (1982). In addition, they did not use a counter-balanced design as had a number of earlier Computer Assisted Assessment (CAA) and Automated Projector Assessment System (APAS) researchers.

In addition to the Apple II Plus, Rock and Nolen's (1982) study employed two floppy-disk drives, a thermal printer, Clock/Calendar

Card and a Sony 17-in video color monitor. The 36 figures of the CPM were programmed into the computer using Applesoft Basic and an Apple graphics tablet. A linking program presented the 36 items and maintained timing, test scoring, and reporting subroutines.

The subjects responded to the CAA presentation of the CPM by selecting "... each item by a color-coded key response matched to a color box placed above each response option displayed on the monitor" (Rock & Nolen, 1982, p. 41). Thus, this study used a customized keyboard as the format of the human-machine interface.

Rock & Nolen's (1982) study had partially demonstrated the same directional relationship that the other automated testing research had, i.e. the automated version when scored with manual norms was found to be more difficult, than the original test (Elwood, 1969, 1972a, 1972b; Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982). This was the finding on the first twelve items (Subtest A) of the thirty-six item CPM. The mean score for the experimental (CAA) version of the CPM was 11.9, while the normative version (Raven, 1977) was 18.32. The Mann-Whitney $U = 58.5$, $p < .05$) thus

showed a significant difference. This assumption regarding the directional relationship based on Rock & Nolen's research is dubious, however. The other two subtest scores did not demonstrate the directional relationship which had heretofore consistently (although not always to a significant degree) been found in other studies. The automated version of the CPM subtests Ab and B (the remaining twenty-four items of the CPM) actually proved to be less difficult than the normative study group's mean scores.

"The mean ranks for subtest Ab were 16.1 (experimental) and 13.82 (normative) (Mann-Whitney U = 88.5, $p > .05$). Mean ranks for subscale B were experimental 15.2 and normative 14.67 (Mann-Whitney U = 102.0, $p > .05$)" (Rock & Nolen p.42)

Thus, unlike the other aforementioned studies, Rock & Nolen (1982) did not demonstrate the consistent, directionally based finding that suggested (generally not to a significant degree) that the APAS presentation of the various tests' items increased the degree of difficulty of the task, thereby lowering the mean scores of the

automated version relative to those of the manually administered test.

As previously discussed, numerous qualifications need to be made prior to generalizing Rock & Nolen's results. In general, their study failed to use a counter-balanced, test-retest design as did others (Overton & Scott, 1972; Knights, Richardson & McNarry 1973; Calvert and Waterfall, 1982; and Watts et al., 1982). The sample size of their study was quite small ($N = 15$), thereby reducing the possibility for the discovery of significant relationships. In addition, they failed to use the recommended two component response format or the recommended feedback wherein the subject's initial (and tentative) response was displayed with opportunities provided for the subject to change his response.

Beyond these qualifications, it may be suggested that the enhanced capabilities of the CAA format of automated testing could have been a factor in the findings which, for the first time within the reviewed literature, suggested that automated testing did not induce greater difficulty than the manual version. Clearly as Rock & Nolen (1982) asserted: "these findings suggest that the

potential utility of computerized testing applied to psychological and educational assessment warrants continuing research investigation" (p.42).

Consequently, in investigating what computer system to employ in the present study, a concerted effort was made to find one which offered the highest degree of graphic resolution and the best means of human to machine interface. It appeared that the putative increased difficulty which the research review had suggested may have been largely due to the relatively poor graphic capabilities of the earlier hardware and low degree of machine-human interaction which provided insufficient opportunities for subjects to learn from the CAA version of the test when it was administered prior to the manual version in a pre-post research design.

The Apple II Plus, which Rock and Nolen (1982) used, had been superseded by the Apple IIe, at the time of this study's initial development during the fall of 1984. The Apple IIe consequently was closely examined during the initial phase of this hardware selection period. The Apple IIe was a very favorable choice, given the preliminary experimental design requirement which necessitated

using a machine in each of the schools where the CAA evaluations were planned to be performed. The school division which authorized this research had embarked on a purchasing plan which had placed at least one Apple IIe in each elementary school within the school division. This factor offered considerable promise, as once software development of the RSPM had occurred, using it within the various schools for running subjects would be relatively uncomplicated. Use of the available computer systems within each of the selected schools would have diminished the design difficulties of hardware transportation, equipment reliability and acquisition, and would have allowed subjects to be tested in more than one school on the same day.

Upon further exploration of the graphic capability of the IIe, it was determined that many of the geometric figures of the RSPM would have a pronounced sawtoothed quality on any diagonal lines or curved figures due to the relatively coarse graphic resolution of the Apple IIe. An additional concern with the Apple IIe centered on the amount of storage space of the graphic figures. As previously mentioned, Rock and Nolen's study automated the 36 figure CPM. The

60 item RSPM, a better suited test for older subjects, would require a greater amount of storage space for the more complicated and more numerous RSPM figures. The use of more than two floppy-disk drives or a hard-disk drive was considered, but determined to be unfeasible due to the added complexity inherent in such customized hardware systems. An a priori goal of this research was to develop a screening device usable by practicing school psychologists. Thus, the complexity of setting up, and using the device had to be kept to a minimum.

The Apple IIe would also have required the use of a keyboard form of interaction, unless the test developer engaged in considerable programming modifications to allow the use of a mouse. Since the mouse-mediated response format had been added to the Apple II series of computers after the development of the machines, no form of highly accurate measurement of motoric output could be readily developed due to the lack of a bit-mapping screen which would allow the pixel-by-pixel determination of mouse (cursor) positioning.

In addition, the RSPM does not require a color output as did the

Raven Coloured Progressive Matrices (CPM) and therefore one of the major attributes of the Apple IIe was not needed.

The Apple Macintosh was introduced in January, 1984. It appeared to offer solutions to all of the aforementioned hardware limitations of the Apple IIe. The larger amount of disk storage on each floppy-disk (400 Kilobytes versus 128 Kilobytes for the Apple IIe) appeared to overcome the greater storage demands inherent in automating the RSPM.

The mouse-mediated format, which was an inherent design consideration of the Macintosh around which all of its software is created, offered numerous advantages. It allowed an extremely easy to understand and use human-machine interface. It allowed the automated RSPM to be developed without requiring the subjects to use a keyboard. This avoided the possible confounding of the test results stemming from distracting effects or uncontrolled effects stemming from subjects' varying degrees of previous exposure to keyboards. In addition, use of a computer system which did not require the custom-building of a special keyboard for the subjects to interact with increased the feasibility of the development of the CAA device.

Another promising aspect in using the Macintosh's mouse was the possibility for the development of an actual behavioral index of a subject's problem solving style through measuring the linear movement of the mouse-mediated cursor across the screen as the subject responded to each item. This measure of 'motoric efficiency' appeared to offer considerable promise as a psychometric parameter.

Use of the mouse-mediated response format additionally allowed the development of the automated RSPM to proceed wherein the subject would only have to essentially make a pointing response in order to interact with the CAA device. Thus, the recommendations of Raven for a two component response format and the inclusion of feedback clearly designating the tentative item chosen by the subject could be implemented in an easy to demonstrate, easy to use, testing format. In addition, while the subject was making his responses on each item, the computer could covertly time his responses and note the degree of purposefulness in which the subject executed his choices on each test item.

During the initial planning phase of this study during the Fall of 1984, Apple Computer Company introduced the 512 Kilobyte version

of the Macintosh. The introduction of this so-called "Fat-Mac" offered this study a very capable machine. It provided a 512 Kilobyte upgrade of the original machine which had 128 Kilobytes of internal random access memory (RAM). This study initially considered the use of two floppy disk drives, one internal and one external, each providing 400 Kilobytes of memory storage per disk. However, as preliminary development of the CAA device occurred, it was determined that 800 Kilobytes of storage did not allow an adequate amount of storage for the CAA software or provide the ease of administration desired. Consequently, an internally mounted, 10 Megabyte (Ten Million Characters of Storage) hard-disk drive was employed. The use of such a device had earlier been ruled out due to cost and portability factors. However, a compact, relatively inexpensive hard-disk drive called a Hyperdrive became available during the development of this study's CAA instrumentation. This hard disk drive actually fits inside the Macintosh computer. It allows the machine to be very easy to use, since no floppy-disks are required. Users of the CAA software mounted inside a Macintosh are only required to turn on the machine and click on the RSPM icon provided by the CAA software program.

The 32-bit processing format of the Macintosh allowed for relatively compact storage of the RSPM items (approximately 13 Kilobytes per figure). The relatively (compared to most microcomputers) large 512 Kilobyte internal memory of the Macintosh computer allowed the computer to hold a very complex CAA program in Random Access Memory, thereby allowing the CAA program to run virtually instantaneously, with the exception of unpacking the graphic images. This additional processing speed was hoped to minimize any differences between the time of administration of the paper and pencil versions of the RSPM and the CAA format of the test. Unfortunately, as discussed in the forthcoming section - **Developing a means of rapidly unpacking the RSPM images and displaying them on the computer screen** - the total test time did vary significantly between the CAA and Manual versions of the RSPM.

The bit-mapping method of graphic presentation utilized by the Macintosh allowed subjects to respond to the automated RSPM through the use of a device commonly referred to as a "mouse". The subject responded simply by rolling the "mouse" device on the table near the

computer screen. A pointer or cursor on the computer screen moved in a manner corresponding to the direction and extent of the "mouse's" movement. The graphic display of 512 x 342 pixels per inch provided very detailed, high resolution images in black, varying shades of gray, and white. The bit-mapping format allowed the Macintosh to be programmed to count the linear pixel by pixel movement of the mouse yielding the previously discussed measure of 'motoric efficiency'. As the reader will note, the high degree of resolution provided very accurate measurement of the subject's pointer movement, in excess of 80 units per inch of movement.

The Macintosh also offered mainframe compatibility. Eventually, highly detailed, item-by-item statistical analysis or more advanced future CAA development will be possible given the Macintosh's 32-bit design which allows for direct, microcomputer to mainframe communication.

Due to the Macintosh's relative infancy, the number of skilled computer programmer's capable of effectively programming it was found to be very limited. In addition, the Macintosh's very advanced, graphically intensive means of displaying information posed

additional difficulties. The very detailed graphic capabilities required very complex instructions. Many of these instructions were referred to as "ROM commands" or commands which directly addressed the Read Only Memory (ROM) circuitry within the computer. The Macintosh had approximately 900 different ROM commands, of which a programmer must have a thorough understanding prior to creating a novel repertoire of computer commands as was required in automating the RSPM. These 900 ROM commands are described within an Eight Hundred page book called Inside Macintosh. (Apple, 1985).

In developing the CAA program using a Macintosh 512 Kilobyte computer, the essential developmental elements were determined to be:

1. Determining the most suitable and computer compatible manner in which to reproduce the RSPM images into electronic media.
2. Developing a means of rapidly unpacking the RSPM images and displaying them on the screen.
3. Developing a means of accurately timing and storing the three component measure of the subjects' response time.

4. Developing a means of counting and storing the pixel-by-pixel movements of the mouse-mediated cursor as the subject completed each of the items of the automated RSPM.
5. Developing a means of noting the subjects' non-fatal errors and recording them.
6. Developing a means of noting the subject's final answer and scoring it.
7. Developing a means of producing cumulative totals of all of the above psychometric data and determining decision rules of how to best score these variables.
8. Determining how to initially present the CAA test format to subject's so that they could grasp it, and yet not confound direct comparisons with the standardized paper and pencil RSPM.
9. Creating directions for examiners of the prototypic device which were clear and concise and allowed a standardized form of presentation.

These CAA development issues will be considered in detail, point by point:

1. Determining the most suitable and computer compatible manner in which to reproduce the RSPM images into electronic media. The Apple Macintosh is an extremely graphically oriented machine. Two drawing programs were considered to reproduce the RSPM images. These were Macdraw and Macpaint. The Macdraw program allowed the use of precise scaling techniques, as X and Y coordinates of the drawing instrument are provided by the computer software. It was recognized that the screen images of the RSPM would have to be approximately 85% the size of the actual standardized manual RSPM images in order to fit on the screen. Thus the use of this program appeared quite promising.

In examining the format with which the ROM based Quickdraw routine of graphic presentation operated, which underlies the presentation of all graphic images on the Macintosh, it was learned that Macpaint images would be required for the computer to readily read and display the designs within the CAA program. Thus, the

hardware requirements of the chosen computer dictated the means of drawing the figures.

The Quickdraw master routine of graphic presentation also required precisely defined rectangles in which to read and execute programmed instructions. Each of the the RSPM images had to be displayed in precisely the same position on the computer screen in order for the CAA program to note the subject's response, highlight the chosen image and provide the programmed prompts which allowed the subject to lock in his response. Consequently two highly accurate templates were draw using the Macdraw program.

In drawing the templates, the RSPM figures had to be positioned in a different manner than the original RSPM. This was because the manual version of the test is presented in a paper booklet approximately 8 1/2 inches wide by 9 1/2 inches in height. The usable computer screen area is approximately 7 inches wide by 5 inches high, thus having the longer axis in the opposite direction as the paper version of the RSPM. Consequently, the placement of the images had to be changed from that of the paper version, which had the large pattern in the uppermost portion of the page, and the 6 or 8

response choices (6 choices in sets A & B; and 8 choices in sets C, D & E) aligned in two rows at the bottom of the page. The means of obtaining the largest scaling of the automated RSPM images (85% of the size of the original paper figures) was provided by fitting the large figure against the upper right portion of the screen and orienting response choices down the left side and along the bottom of the large figure. The response choices were numbered as in the paper version of the test, although the subjects did not have to use the numbers to indicate their answers.

Two, roughly 85% scale, templates were made of the overall RSPM example pattern and the 6 or 8 test items. These templates were created using Macdraw and then converted into Macpaint images. The actual RSPM figures were then drawn by starting with the respective empty template for sets A & B or sets C, D & E and generating the correct patterns for each item using the Macpaint graphics application.

The numerous graphic tools of the Macpaint application allowed relatively accurate, visually pleasing reproductions of the original RSPM figures. Every effort was made to comply with the details of

the original images. The "fatbits" tool within the Macpaint application allowed the pixel by pixel drawing of the images, thereby allowing detail in excess of 1/80th of an inch.

An example of the Macpaint generated images are shown in Appendix C. The Macpaint graphics program allows for the printing of images generated within its program. The CAA RSPM program unfortunately does not provide the capability to perform a "screen dump" to the printer, which would allow for the previously mentioned prompt lines of this program to be shown, as they appear, during the test administration. As the reader may note, the images are very close reproductions to the original published RSPM images and the 85% scaling did not appear to diminish discriminability. It should be noted that the actual CAA figures were shown via transmitted light from the computer's Cathode ray tube and therefore had sharper contrast than that afforded by the reflected light of the printed material.

2. Developing a means of rapidly unpacking the RSPM images and displaying them on the screen. As previously

mentioned each RSPM Macpaint image required an average of 13 Kilobytes of storage space. However the 512 x 342 pixel screen of the Macintosh computer contains approximately 175 Kilobytes of graphic imagery ($512 \times 342 = 175,104$ pixels). Thus, the Macpaint program provides a high degree of compression when it "packs" the 175 Kilobytes of screen information into the average of 13 Kilobytes of stored information.

The reverse of the compression or "packing" process is the screen depiction or "unpacking" of the stored image. Both processes require a considerable amount of computer and storage medium processing time. No time is required during actual CAA presentation of the images for packing, since the images are already stored on the storage medium or disk. However, to display these images, they must be read by the computer and unpacked into the detailed screen image and displayed on the screen.

In testing this process using the standard floppy disk drives of the Macintosh (an internal and external drive) it was determined that 4 to 7 seconds was required to unpack each image. In addition it was determined that the single Central Processing Unit (CPU) format of

the Macintosh only allowed it to perform one task at a time. Thus, when the CPU was unpacking a RSPM image, it could not accurately monitor response time or motoric efficiency. Therefore, the CAA requirement of monitoring response times and motoric efficiency determined that the program must unpack the RSPM images after each response, rather than during the previous response. This meant that subjects would have to view a blank screen for 4 to 7 seconds while the computer read the packed information off the floppy disk.

Further investigation indicated that a hard disk drive, which operates at approximately 10 times the speed of the floppy disks, would reduce the "read time" of each RSPM image to less than one second. This time, combined with the CPU's unpacking time, would allow the next sequential RSPM image to appear after approximately a one second delay. This time delay appeared acceptable, while the 4 to 7 second delay imposed by using the floppy disk storage medium appeared unacceptable.

As indicated earlier, the use of a hard disk drive in this study had been ruled out initially, due to added complexity, cost, and diminished portability. Fortunately, the General Computer Company

introduced the Hyperdrive™ 10 Megabyte hard disk drive in February, 1985. This hard disk drive system mounted inside the Macintosh and actually decreased the complexity of operating the computer. This diminished complexity stemmed from the ability of the Hyperdrive™ to open directly to the CAA program upon being turned on. Thus, the psychologists participating in the study had only to be taught to turn on the machine, click on the CAA icon which started the program, and administer the standardized instructions to the subject.

3. Developing a means of accurately timing and storing the three component measure of the subjects' response time. The Macintosh computer has an integral unit of time, called a "tick" which is equivalent to 1/60th of a second. Numerous ROM based commands can address the tick unit and thereby perform certain tasks after a pre-assigned number of ticks. In developing the response time measure, it was determined that this degree of accuracy was sufficient for monitoring human responses.

The Initial Response Time measure was developed in an attempt to measure the impulsivity of subjects' responses. A small, invisible

rectangle approximately one inch square was defined around the box shown on the screen for locking in responses. After the completion each response, the subject had to place his cursor within the approximately 1/2 inch square box on the screen and click his mouse in order to move on to the next RSPM item. Thus, as the screen refreshed itself, the mouse-mediated cursor would always appear in the same place on the screen until the subject moved the mouse. It was believed that since the cursor was programmed to disappear after locking in the previous response, and since no other stimuli was present, mouse movement outside the pre-defined, invisible rectangle would not occur until the screen refreshed itself with the new image, except in the case of impulsive individuals. The results of this study indicated that this hypothesis was too simplistic, as no significance was shown for Initial Response Time in any of the statistical tests performed as part of this study. It appears that subjects may have moved their mouse to the area of the choices, in preparation for the next item.

Intermediate Response time was defined as the time from when the subject moved the mouse outside of the invisible, pre-defined

rectangle in the upper right-hand corner of the screen (where the subject had locked in his previous response) until he clicked his mouse on his first choice of an answer for the subsequent item.

Final Response time was defined as the time from when the subject clicked on his first choice of an answer within a RSPM item, until he locked in his response. Thus if he changed his answer or went off-task after tentatively choosing an answer, this time elapsed would be recorded as his Final Response time.

All three of these response time measures were measured in tick unit. This unit of measure can be readily converted to milliseconds by multiplying the tick units by $16 \frac{2}{3}$. Seconds may be derived by dividing tick units by 60.

4. Developing a means of counting and storing the pixel-by-pixel movement of the mouse-mediated cursor as the subject completed each of the items of the automated RSPM. The measure of motoric efficiency also used the tick unit of time. The CAA program was designed so that at each tick (every

1/60th of a second) the computer would "look" and note the X and Y coordinates of the mouse-mediated cursor's position. At the next tick, the new X and Y coordinates were noted and the linear distance between these two positions was computed. A running total was maintained of these linear distances until the subject clicked his mouse on the first chosen answer of each item. At this time, the CAA program instructed the computer to store this linear distance as the Phase One Motoric Efficiency value for that item. The same process occurred during the Phase Two Motoric Efficiency measure, which measured the linear distance of the mouse-mediated cursor from the initial tentatively chosen item until the subject locked in his response and thereby moved on to the next item.

5. Developing a means of noting the subjects' non-fatal errors and recording them. Non-fatal errors were defined as those responses which a subject made but then changed, prior to locking them in, so that they were not scored in the RSPM total number correct. A response was defined as a mouse click. Thus, when a subject clicked on an RSPM answer, the number of his choice

was recorded. In the event that the subject clicked the mouse outside of a sensitized rectangle near an answer, a zero (0) was recorded for this response. The final response a subject made and subsequently locked in was scored. If this final response was incorrect it was scored as a fatal error.

An item by item tally of the subject's responses was kept. The CAA program wrote the subject's choice onto the hard disk each time he clicked the mouse.

6. Developing a means of noting the subject's final answer and scoring it. As described above, the subject's final scored response was the one he locked in. Upon locking in the subject's response, the CAA program recorded his choice, compared it with the encoded scoring key and noted whether the answer was correct or not. At the completion of the test, a tally of correct answers was recorded as part of each student's summary information.

7. Developing a means of producing cumulative totals of all of the above psychometric data, and determining decision rules of how to best score these variables. For each item administered the computer recorded the three response time measures, two motoric efficiency measures and the non-fatal and fatal errors for that item. For a given subject, these scores were totalled across all items according to the respective parameter categories. The totaling process took into account the varying number of repetitions which may have occurred for items A_1 and A_2 which both the standardized manual and automated CAA versions of the RSPM allowed to be repeated until the subject understood the nature of the task.

8. Determining how to initially present the CAA test format to subject's so they could grasp it, and yet not confound direct comparisons with the standardized paper and pencil RSPM. This study was initially proposed to provide a form of error-free learning prior to beginning the CAA test. It was proposed that the CAA program would provide practice items with

very overt correct responses which gradually would become more difficult (stimulus fading). If the subject made incorrect responses on these practice items, the CAA program would have interacted with the subject, inducing them to correct their responses. It was determined that such extensive directions might confound the comparisons between the CAA format and the manual version of the test. Therefore a form of directions as close to the manual version's as possible were developed.

To allow for initial practice and familiarization with the computer items A_1 and A_2 of the CAA program was designed to repeat if the examiner depressed the "Caps Lock" key on the computer keyboard. Thus, if the examiner determined that the subject did not understand the nature of the task, the examiner could depress the "Caps Lock" key and items A_1 and A_2 would repeat as many times as necessary, providing more opportunity for the subject to learn how to use the CAA device. The ability to repeat items A_1 and A_2 is provided

within the standardized directions of the manual RSPM, and thus was included in the automated version as well.

It must be emphasized that only during the initial instruction phase of the automated RSPM that the keyboard is needed. During this instructional phase the keyboard is used exclusively by the examiner while the subject is becoming familiar with the mouse.

9. Creating directions for examiners of the prototypic device which were clear and concise and allowed a standardized form of presentation. Directions for the CAA device were as similar to the manual version of the RSPM as possible.

A major difference was the necessity of teaching the subjects to use the mouse to indicate their responses. On the manual version of the test children are already familiar with using pencil and paper to record their answers.

The standardized directions used by the examiners in this study are shown below:

Standardized Instructions for the

Computerized Raven

Turn on the computer with switch on back of machine (left side). The keyboard should be plugged into the front of the machine. The screen will automatically display the files which are available

Use the Mouse to move the arrow to the **"Automated Raven"** display and click twice on the Mouse button to open the file. You will be asked for a password. Using the keyboard, type in the password: **mac** (in lower case) and hit the Return key. Answer the next questions by simply typing in the information and hitting the Return key after each answer. If you make a mistake, just use the backspace key to erase and then retype your answer. The first test item will automatically appear when you have answered all the questions.

Have the student sit in front of the computer. Show the child the Mouse and say:

"You can use this Mouse to move the arrow and point to objects on the screen (demonstrate moving the arrow and then point to the main figure). See this figure? It is a pattern that has a piece missing from it. These pieces here (move arrow across each possible answer) are all the right shape to fit in the space but only one will complete the pattern."

Move the arrow to *1 and click. Explain why it is not the right answer. Then click on *2 and then *3 and explain why they are not the right answers either. Then click on *6 and explain how it is almost correct. Now ask the child to:

"Use the mouse to point to the answer which is correct, and then click the button on the mouse."

Explain how clicking the mouse button on an answer makes a black border appear around that piece and also creates a black box in the corner of the screen. Show the child that he can change his answer by moving the arrow to a different piece and clicking the button. Explain that when he is satisfied with his answer, he can **"lock it in"** by moving the arrow to the box in the upper right corner and clicking the button. Now allow the child to choose the correct answer, click on it, and lock it in. The next item will automatically appear.

When the second screen has appeared, press down the Caps Lock Key.

Explain that **"Each screen will have a pattern with a piece left out."**

Say: **"All you have to do each time is point to the piece which is correct to complete the pattern and click the Mouse. The items are simple at the beginning and get harder as you go on. If you pay attention to the way the easy ones go, you'll find the later ones less difficult. Just point to the correct piece with the Mouse, click on it, and lock in your answer by pointing to the box in the corner and clicking the Mouse button."**

Depress the Caps Lock key (left of keyboard). Let the child complete the second item if he can. Give further explanation as needed. The screen will now revert to the first item again. Release the Caps Lock key. Give extra explanation as required. After understanding of the task is demonstrated on items one and two say:

"Go ahead at your own pace. See how many you can get right. You can have as much time as you like. There is no need to hurry. Be careful. Remember,

each time only one piece is correct. When you have finished you will see a stop sign...(and examiner may instruct the child what their next action should be, to meet examiners situation.)"

Observe child do #1 and #2 again. If child has difficulty, depress the **Caps Lock Key** again (to repeat items #1 and #2 again) and re-explain the directions to the child until he understands the task. If the child demonstrates understanding of items #1 and #2, allow him to proceed with no further assistance. **(The Caps Lock Key must be released for the test to continue.)** When the child is ready to proceed with the automated test, unplug the keyboard from the front of the machine and set it aside. No supervision of the child's behavior is required, provided he is capable of remaining on task. The test generally requires 18-31 minutes. You may find working in a nearby room, where the child inform you of completion of the test, very convenient.

Encourage the child to attempt all 60 items. If he refuses or gives up (a very rare occurrence), just plug the keyboard back into the machine and depress the Return Key to terminate the test. If the child completes the test and gets a Stop Sign, just plug in the keyboard and hit the **Return Key**. The screen will now revert to the original display of available files. You may switch the machine off (switch is on back of machine) or double click on the **"Automated Raven"** symbol to begin testing another student.

Thats all there is to it! Thanks for your assistance.

Summary: This chapter has described the major considerations which were made in the development of the CAA version of the RSPM. A major conclusion that may be drawn is that psychologists willing to learn some technical computer terminology and principles of computer operation may now develop assessment devices which demonstrate comparable reliability and validity as paper and pencil tests. Furthermore, these newly developed CAA instruments may offer a greater degree of psychometric capability than was previously attainable to paper and pencil assessment instruments.

CHAPTER V

RESULTS

Due to the numerous hypotheses and the multivariant forms of analysis performed to test some of them, the presentation of the results will vary from the prescribed manner of presentation. Discussion of the findings will follow the data presentation for each tested hypothesis in an effort to diminish the need for replication of the details necessary to supplement the discussion of these research findings.

Throughout all tests of hypotheses, the pre-assigned Alpha level is .05. Relationships will be considered significant if they fall at or below this level of significance. The actual measured level of significance will be reported for readers interested in the degree of measured significance.

HYPOTHESIS ONE:

1. There will be a non-significant difference between the computer assisted administration (CAA) of the RSPM and the manual paper and pencil (PP) administration of the the same test when compared in a counter-balanced, test-retest design.

Statistical test employed: Tukey's Studentized Range -HSD (Honest Significant Difference) tests were performed on the CAA then PP group (designated CP) and on the PP then CAA group (designated PC). Tukey's Studentized Range (HSD) test was performed as a second order analysis to the Multiple Analysis of Variance (MANOVA SPSS program). The results shown below are from the MANOVA performed on the data set treating test order (CP vs. PP) as an independent variable. A MANOVA was also performed in which test order was treated as a dependent variable which yielded slightly different numeric results but which concurred with the interpretation of these findings (see Hypothese Three).

<u>GROUPING</u>	<u>(N)</u>	<u>PRE-TEST MEAN(M)</u>	<u>POST-TEST M</u>
CP	21	40.048	42.952
PC	23	40.826	42.217

MINIMUM SIGNIFICANT DIFFERENCE

Pre-test Tukey = 5.90397

Post-test Tukey = 5.30756

FINDINGS: No significant difference found between pre-test or post-test mean scores.

Discussion of Hypothesis One's Findings: The findings shown above support this study's hypothesis that CAA testing, which was developed to provide an adequate degree of interactive feedback of the subjects' responses and required a two-component response which allowed for the subjects to change their responses, produced results comparable to those of the paper and pencil format of the test instrument.

These findings further suggest that the CAA format of testing (as opposed to the apparently less adequate Automated Projector Assessment Systems (APAS) allowed for Automated tests which have the same degree of difficulty (as opposed to the generally found

indications that automated tests were more difficult) as the pencil and paper tests. This finding was initially found in Rock and Nolen (1982) in a confounded manner, where the first twelve items of the test (set A) demonstrated a significantly greater degree of difficulty when compared with the paper and pencil normative group data, while the remaining twenty-four items (sets Ab and B) were found to yield mean scores which suggested that the CAA version of the test was no more difficult than the paper and pencil version.

This study's findings of comparable mean scores between the CP and PC group scores further support the previously cited research findings which suggest that CAA test formats could yield results equal to those of the manual test format.

HYPOTHESIS TWO:

2. The computer assisted administration (CAA) of the RSPM will demonstrate a non-significant difference between pre-test and post-test scores (using an appropriate mean comparison).

Statistical test employed: Tukey's Studentized Range - HSD (Honest Significant Difference) tests were performed on the CAA then CAA group. Tukey's Studentized Range (HSD) test was performed as a second order analysis to the Multiple Analysis of Variance (MANOVA SPSS program). The results shown below are from the MANOVA performed on the data set treating test order (CAA vs. CAA) as an independent variable.

<u>GROUPING</u>	<u>(N)</u>	<u>PRE-TEST</u>		<u>POST-TEST</u>	
		Mean	Stn. Dev.	Mean	Stn. Dev.
CAA *CORRECT	21	39.667	9.232	42.143	8.890
CAA DERIVED IQ	21	103.095	14.352	107.333	15.143

MINIMUM SIGNIFICANT DIFFERENCE

Pre-test * Correct Tukey = 5.90397

Post-test * Correct Tukey = 5.30756

FINDINGS: No significant difference was found between pre-test or post-test mean score. An additional T-test was performed on the results and no significant difference was found between pre-test and post-test means. The mean derived RSPM IQ scores are shown to provide a standard score depiction of the results. As may be seen, the relationship between Pretest and Posttest IQ scores is very close. Discussion regarding the creation of the RSPM IQ score is contained in the forthcoming Hypothesis Three's - Statistical test employed section.

Discussion of Hypothesis Two's Findings: The test-retest measure of the CAA RSPM indicates, through a mean comparison, that this test instrument has a high degree of reliability, when used as a repeated measure over a two-week period. Generalization of the reliability of the instrument over a longer time period appears promising. However, more extensive reliability assessment of the test appears advisable.

HYPOTHESIS THREE:

3. The computer assisted administration of the RSPM will demonstrate a statistically acceptable degree of reliability when measured in a test-retest format [using a reliability coefficient].

Statistical test employed: A correlational matrix was generated in which all group parameters of the automated RSPM were compared with each other. These were: Total test score, Total test IQ Score, Initial Response Time; Intermediate Response Time; Final Response Time; Phase One Motoric Efficiency; Phase Two Motoric Efficiency and Total Test Time. Both the Pre-test and Post-test measures of each parameter were compared, thus making for sixteen (16) rows and columns of comparisons. As may be seen from examining the correlation table, shown on the three pages following this Hypothesis' discussion, there are numerous correlations which suggest interesting relationships among the various automated RSPM test parameters. These will be examined in forthcoming discussions of other hypotheses.

As may be seen a Total Pre-test IQ and Post-test IQ of the RSPM was generated. It was necessary to convert the percentile scores of the standardized paper and pencil RSPM to a standard score to allow for linear comparisons. In addition to attempting to yield more exact placement of subjects scores in making linear comparisons with WISC-R IQ scores, IQ scores provided a means to control the influence of age factors between groups. These IQ scores were generated by using the RSPM 1979 Smoothed Summary Norms (Raven, 1983) which presented the test data in percentile scores, providing 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentile scoring positions for each 6 month age range. These respective standardized RSPM percentile score positions at times left 'holes' in the data table of up to 6 correct responses on the RSPM on which the scorer was required to round down to the previous response score that was shown on the table. This scoring mechanism was found to greatly diminish the accuracy of score comparisons between various groups of subjects.

The interpolated IQ score table is shown in Appendix A. Recent normalization of the RSPM has been performed in the United States

The aforementioned 1979 Summary Norms were based "on a nationally representative sample of British schoolchildren, excluding those attending special schools" (Raven, 1983, p. SPM21). This 1979 norm table is reported by John Raven to concur very closely with the recently completed United States normative sample. (Raven, 1985)

MEASURED TEST-RETEST CORRELATION: $r = .87606$

FINDINGS: The intercorrelations between the Pre-test and Post-test mean Total Score is quite high and concurs very well with the test-retest reliability reported in the literature.

Discussion of Hypothesis Three's Findings: The favorable results regarding the CAA RSPM's reliability suggests that use of the standardized norms obtained from the 1979 Summary Norms of British schoolchildren provides an adequate means of interpreting the CAA version of the RSPM. Raven (1985) has indicated that the recently completed United States standardization is quite compatible with the 1979 British norms.

HYPOTHESIS 3 - 1

	TOPRE	TOPOS	PREIQ	POIQ	PRET1	PRET2	PRET3
TOTAL PRE-TEST SCORE (TOPRE)	1.00000	0.87606	0.92060	0.74291	-0.30171	0.58373	-0.63859
TOTAL POST-TEST (TOPOS)	0.87606	1.00000	0.83068	0.92665	-0.29756	0.58161	-0.69184
DERIVED IQ PRE-TEST (PREIQ)	0.92060	0.83068	1.00000	1.00000	-0.23995	0.58553	-0.45279
DERIVED IQ POST-TEST (POSIQ)	0.74291	0.92665	0.82065	0.92665	-0.22563	0.53203	-0.53799
PRE RESPONSE TIME 1 (PRET1)	-0.30171	-0.29756	-0.23995	-0.22563	1.00000	-0.30424	0.21403
PRE RESPONSE TIME 2 (PRET2)	0.58373	0.58161	0.58553	0.53203	-0.30424	1.00000	-0.15453
PRE RESPONSE TIME 3 (PRET3)	-0.63859	-0.69184	-0.45279	-0.53799	0.21403	-0.15453	1.00000
POST RESPONSE TIME 1 (POSTT1)	0.00812	0.01419	-0.00938	-0.00938	0.56119	0.28970	0.09377
POST RESPONSE TIME 2 (POSTT2)	0.27807	0.60001	0.39179	0.65805	-0.22029	0.46845	-0.29742
POST RESPONSE TIME 3 (POSTT3)	-0.14355	-0.05265	-0.14574	-0.09820	0.04854	-0.11059	-0.00400
PRE MOTOR EFFICIENCY 1 (PREME1)	0.32870	0.31686	0.44318	0.38859	-0.57982	0.36222	-0.06987
PRE MOTOR EFFICIENCY 2 (PREME2)	-0.29022	-0.41115	-0.22446	-0.36352	-0.34523	-0.36293	0.42757
POST MOTOR EFFICIENCY 1 (POSTME1)	0.19173	0.27171	0.29615	0.36526	-0.48229	0.02908	-0.07247
POST MOTOR EFFICIENCY 2 (POSTME2)	-0.01789	-0.15376	-0.09369	-0.26047	-0.13816	-0.45611	-0.23120
PRE -TEST TIME (TOTIMPR)	0.24938	0.23974	0.30050	0.23407	0.32415	0.76198	0.18353
POST-TEST TIME (TOTIMPO)	0.17911	0.46778	0.25765	0.48210	0.05355	0.46144	-0.20015

HYPOTHESIS 3 - 2

	POSRT1	POSRT2	POSRT3	PREME1	PREME2	POSME1	POSME2
TOTAL PRE-TEST SCORE (TOPRE)	0.00812	0.27807	-0.14355	0.32870	-0.29022	0.19173	-0.01789
TOTAL POST-TEST (TOTPOS)	0.01419	0.60001	-0.05265	0.31686	-0.41115	0.27171	-0.15376
DERIVED IQ PRE-TEST (PREIQ)	-0.00938	0.39179	-0.14574	0.44318	-0.22446	0.29615	-0.09369
DERIVED IQ POST-TEST (POSIO)	-0.03785	0.65805	-0.09820	0.38859	-0.36352	0.36526	-0.26047
PRE RESPONSE TIME 1 (PRERT1)	0.56119	-0.22029	0.04854	-0.57982	-0.34523	-0.48229	-0.13816
PRE RESPONSE TIME 2 (PRERT2)	0.28970	0.46845	-0.11059	0.36222	-0.36293	0.02908	-0.45611
PRE RESPONSE TIME 3 (PRERT3)	0.09377	-0.29742	-0.00400	-0.06987	0.42757	-0.07247	-0.23120
POST RESPONSE TIME 1 (POSRT1)	1.00000	0.01690	0.04665	-0.45769	-0.54355	-0.64192	-0.31453
POST RESPONSE TIME 2 (POSRT2)	0.01690	1.00000	0.27856	0.38807	-0.19402	0.48849	-0.21402
POST RESPONSE TIME 3 (POSRT3)	0.04665	0.27856	1.00000	0.12707	0.22109	0.29834	0.63904
PRE MOTOR EFFICIENCY 1 (PREME1)	-0.45769	0.38807	0.12707	1.00000	0.40747	0.79444	0.12751
PRE MOTOR EFFICIENCY 2 (PREME2)	-0.54355	-0.19402	0.22109	0.40747	1.00000	0.55004	0.49382
POST MOTOR EFFICIENCY 1 (POSME1)	-0.64192	0.48849	0.29834	0.79444	0.55004	1.00000	0.27230
POST MOTOR EFFICIENCY 2 (POSME2)	-0.31453	-0.21402	0.63904	0.12751	0.49382	0.27230	1.00000
PRE -TEST TIME (TOTIMPR)	0.67625	0.24057	-0.05053	-0.04252	-0.48547	-0.32507	-0.55499
POST -TEST TIME (TOTIMPO)	0.41465	0.67432	0.51983	0.17221	-0.31214	0.21977	-0.12080

HYPOTHESIS 3 - 3

	TO TIMPRE	TO TIMPO
TOTAL PRE-TEST SCORE (TOPRE)	0.24938	0.17911
TOTAL POST-TEST (TOTPOS)	0.23974	0.46778
DERIVED IQ PRE-TEST (PREIQ)	0.30050	0.25765
DERIVED IQ POST-TEST (POS IQ)	0.23407	0.48210
PRE RESPONSE TIME 1 (PRERT1)	0.32415	0.05355
PRE RESPONSE TIME 2 (PRERT2)	0.76198	0.46144
PRE RESPONSE TIME 3 (PRERT3)	0.18393	-0.20015
POST RESPONSE TIME 1 (POSRT1)	0.67625	0.41465
POST RESPONSE TIME 2 (POSRT2)	0.24057	0.87432
POST RESPONSE TIME 3 (POSRT3)	-0.05153	0.51983
PRE MOTOR EFFICIENCY 1 (PREME1)	-0.04252	0.17221
PRE MOTOR EFFICIENCY 2 (PREME2)	-0.48547	-0.31214
POST MOTOR EFFICIENCY 1 (POSME1)	-0.32507	0.21977
POST MOTOR EFFICIENCY 2 (POSME2)	-0.55499	-0.12080
PRE -TEST TIME (TOTIMPR)	1.00000	0.44729
POST-TEST TIME (TOTIMPO)	0.44729	1.00000

HYPOTHESIS FOUR:

4. The computer assisted administration of the RSPM will not show any significant order effect (CAA - then - Manual versus Manual - then - CAA) in appropriate statistical comparisons.

Statistical test employed: A multiple analysis of variance (MANOVA) was performed using the SPSS statistical program. The results shown below are from the MANOVA performed on the data set treating order of test administration (CP vs. PP and PP vs. CC) as the independent variable. A MANOVA was also performed in which test order was treated as a dependent variable. This treatment yielded slightly different numeric results, but concurred with the interpretation of these findings.

The MANOVA comparisons held the order effect constant treating it as an Independent Variable. The following factors were treated by the MANOVA as dependent variables (DV): Derived IQ Pre-test; Derived IQ Post-test; Age of Subject; Total Pre-test Score; Total Post-test Score; Total Time of Pre-Test; Total Time of Post-test.

Given the above manner of comparisons, obtaining a significant (alpha level = .05) F score would indicate that there is a significant order of test administration effect for the specified dependent variable.

<u>DEPENDENT VARIABLE</u>	<u>F VALUE</u>	<u>SIGNIFICANCE LEVEL</u>
IQ PRE-TEST	0.00	.9974
IQ POST-TEST	0.23	.7949
AGE	0.22	.7995
TOTAL PRE-TEST SCORE	0.12	.8886
TOTAL POST-TEST SCORE	0.08	.9228
TOTAL TIME PRE-TEST	7.45	.0013****
TOTAL TIME POST-TEST	3.09	.0528

**** REPRESENTS SIGNIFICANT RESULTS

FINDINGS: No significant differences were found between pre-test or post-test F scores relating to the learning effect provided by order of test administration. This finding suggests that no disparity exists between the learning effect the CAA test produces when given as a pre-test, versus the corresponding learning effect of the Manual version when it is given as a pretest.

A significant order effect was found for the total time of test variable in the pre-test condition. Consequently a Tukey's Studentized Range (HSD) test was performed, in order to discern the specific order of treatment which produced the significant F score. In addition, a Tukey's Studentized Range - HSD test is shown for the non-significant total time of test in the post-test condition. This table presents the results of the treatment groups in the same order as the pre-test condition; however, the mean time of test values produce a different rank ordering. No significant order effect was found in the post-test condition.

TUKEY'S STUDENTIZED RANGE (HSD) FOR PRE-TEST TOTAL TIME

<u>TEST ORDER</u>	<u>(N)</u>	<u>PRE-TEST TIME MEAN</u>	<u>INTERACTION</u>
CAA -MANUAL	21	31.190 Minutes	YES W/ MAN-CAA
CAA - CAA	21	26.905 Minutes	NO
MANUAL -CAA	23	23.696 Minutes	YES W/ CAA-MAN

TUKEY MINIMUM SIGNIFICANT DIFFERENCE = 4.70498

TUKEY'S STUDENTIZED RANGE (HSD) FOR POST-TEST
TOTAL TIME

<u>TEST ORDER</u>	<u>(N)</u>	<u>POST-TEST TIME MEAN</u>	<u>INTERACTION</u>
CAA -MANUAL	21	21.333 Minutes	NO
CAA - CAA	21	20.714 Minutes	NO
MANUAL-CAA	23	24.043 Minutes	NO

TUKEY MINIMUM SIGNIFICANT DIFFERENCE = 3.47232

Discussion of Hypothesis Four's Findings: As previously stated no significant order effect was observed between the pre-test and post-test scores of the total RSPM test scores or the derived IQ scores. This suggests that the automated version of the RSPM is quite similar to the manual version in the learning opportunities it offers subjects when given as a pre-test.

A significant order effect for the pre-test total time of test was observed. The Tukey Studentized Range - HSD test indicated that the order effect was significant between the CAA then Manual group versus the Manual then CAA group. The CAA then CAA group did not significantly interact with either of these other groups, as this

group's total time of test fell between the other groups' scores on this variable.

The results indicate that the CAA then Manual group spent an average of over 31 minutes on their CAA pre-test. The Manual then CAA group spent an average of approximately 23.7 minutes on their initial exposure to the RSPM, which was given in the manual format.

The lack of a Manual then Manual group in this study makes interpretation of these results uncertain. The results of the Tukey Studentized Range - HSD test on the non-significant post-test total time of test has a different rank ordering of the groups. The Manual then CAA group (which in the pre-test comparison had the shortest test time) had the longest total time of test. The CAA then Manual group had the intermediary time, while the CAA then CAA group had the shortest total post-test time. Initial analysis would suggest that the CAA format required more time than the Manual version of the test. This assumption is confounded however by the CAA then CAA finding which produced the lowest post-test total time of test. The CAA then CAA finding appears logical as this group simply took the same test over. Therefore, it is feasible to assume that the time

required for this group to complete their second exposure to an identical task would be considerably less than their initial time on this task. Unfortunately this assumption does not provide a determination of the extent to which the CAA format required more time in the comparison between the Manual and automated versions of the RSPM.

HYPOTHESIS FIVE:

5. Automated (computerized) administration of the Raven Standard Progressive Matrices (RSPM) will demonstrate a significant relationship in a post hoc manner with the outcome of formal psychoeducational evaluations using the WISC-R.

Statistical tests employed: To test this hypothesis a stepwise multiple regression was performed wherein all the parameters of the automated RSPM were examined to determine their respective degree of loading on the WISC-R verbal, performance and full scale IQ scores of the students in the WISC-R - automated RSPM group (N=20). The automated RSPM scores which were examined were: Total RSPM score; RSPM IQ [using interpolations (as described in Hypothesis Three) of Raven's 1979 norms (Raven, 1983)]; Phase One Motoric Efficiency; Phase Two Motoric Efficiency; Initial, Intermediate and Final Response Times, and Total Test Time.

Each of the WISC-R IQ scores were treated in this Multiple Regression as Dependent Variables. In interpreting these results, the R Square value may be used as an indication of the percentage of

variance accounted for by the parameter which produced a significant loading, eg: the first R Square value of .35088 indicates that this value accounts for 35.088 % of the variance of the WISC-R verbal IQ parameter. The stepwise multiple Regression for each of the Dependent Variables produced the following significant ($\alpha=.05$) loadings:

WISC-R VERBAL IQ LOADINGS

<u>STEP</u>	<u>RSPM VARIABLE</u>	<u>R SQUARE</u>	<u>ESCORE</u>	<u>SIGNIF. LEVEL OF INTERCEPT</u>
1	RSPM IQ	.35088	9.73	.0029
2	MOTOR EFFIC. 2	.48547*	8.02	.0501

WISC-R PERFORMANCE IQ LOADINGS

<u>STEP</u>	<u>RSPM VARIABLE</u>	<u>R SQUARE</u>	<u>ESCORE</u>	<u>SIGNIF. LEVEL OF INTERCEPT</u>
1	RESPONSE TIME 3	.58308	25.17	.0002
2	RSPM IQ	.66189*	16.64	.0291

WISC-R FULL SCALE IQ LOADINGS

<u>STEP</u>	<u>RSPM VARIABLE</u>	<u>R SQUARE</u>	<u>FSCORE</u>	<u>SIGNIF. LEVEL OF INTERCEPT</u>
1	RSPM IQ	.37317	10.72	.0024
2	MOTOR EFFIC.1	.58502*	9.27	.0307

* These values represent the total of the first and second loading.

FINDINGS: The results of the post hoc comparison of the Computerized administration of the RSPM and the previously administered WISC-R scores demonstrated that the CAA obtained RSPM IQ score significantly predicted the WISC-R verbal, performance, and full scale IQ scores, thereby supporting the assertion of Hypothesis Five. The RSPM was the strongest predictor of the WISC-R verbal and full scale scores. The CAA obtained measures of motoric efficiency (previously unobtainable prior to the development of this instrument) parameters also demonstrated significant loadings. Motoric efficiency 2 significantly loaded with the WISC-R Verbal IQ

score and motoric efficiency 1 demonstrated a significant relationship with the WISC-R Full Scale IQ scores. The WISC-R performance IQ score was most strongly predicted by the CAA obtained measure of Final Response Time and also significantly predicted by the automated RSPM IQ score.

Discussion of Hypothesis Five's Findings: The findings of the relationship between the automated RSPM and the WISC-R verbal and full scale IQ scores are consistent with Hypothesis Five's anticipated relationship of the RSPM with the WISC-R. It would appear quite reasonable and likely that one IQ test would demonstrate the ability to significantly predict the results of another instrument also designed to predict academic success. In addition, the best predictor would also appear to be the IQ score of each test toward the other. Thus, this study's significant relationship between the verbal and full scale WISC-R IQ scores and the automated RSPM IQ scores appears to be quite reasonable.

The relationship between the WISC-R performance IQ score and the CAA obtained measure of Final Response Time was unexpected

The RSPM is generally recognized as a test of non-verbal reasoning which therefore would appear to have its highest degree of concurrence with the WISC-R performance IQ scale. This intuitive assumption however is not clearly supported by the research on the RSPM. Raven cites Deutsch, Katz and Jensen (1968) as having "suggested that successful performance on so-called non-verbal tasks of intellectual ability actually requires spontaneous verbalization on the part of the subject" (Raven, 1983, p. SPM 12). Burke and Bingham (1968) found further support for this "Verbalization Hypothesis" (p. SPM 12). They observed that verbal content heavily determined the general ability factor which accounted for most of the RSPM's variance in their study.

Thus the failure of the automated RSPM IQ to demonstrate the strongest relationship to the WISC-R performance IQ score may be due, in part, to the stronger relationship of the RSPM to verbalization factors. In addition, the relation of CAA derived response time in predicting the WISC-R performance IQ had not been previously noted. Specifically, the newly developed CAA obtained measure of Final Response Time appeared to have a very strong relationship to the

WISC-R performance IQ. It would appear that the importance of time, and specifically the speed with which subjects determine that they have accurately completed their RSPM responses, lock-in their responses and move on to the next item had a pronounced and significant relationship with how successfully subjects completed the performance subtests (which are all timed tasks) of the WISC-R.

The finding that the newly developed, CAA obtained measure of motoric efficiency had significant relationships with the WISC-R verbal and full scale IQ scores, indicated that subjects' degree of purposefulness in performing the automated RSPM had a strong relationship with the WISC-R verbal and full scale IQ scores.

HYPOTHESIS SIX:

- 6 The automated RSPM will yield a statistically significant correlation with the WISC-R scores.

Statistical test employed: A correlational matrix was generated in which the automated RSPM's total score and IQ scores were compared with the WISC-R verbal, performance and full scale IQ scores of the WISC-R - automated RSPM group of subjects (N=20). A critical value of $r < .509$ was determined for the Pearson Product Moment Correlation which was used in the reported correlations. Thus correlations which exceeded the critical value of .509 may be considered significant.

AUTOMATED RSPM - WISC-R CONCURRENT VALIDITY
CORRELATIONS

<u>WISC-R IQ</u>	<u>RSPM TOTAL SCORE</u>	<u>RSPM IQ SCORE</u>
VERBAL	.499	.599****
PERFORMANCE	.373	.462
FULL SCALE	.508****	.618****

**** indicates significant correlational relationships.

FINDINGS: The intercorrelations between the WISC-R verbal and full scale IQ's and automated RSPM are statistically significant ($\alpha = .05$). These findings concurred very well with the manual RSPM - WISC-R concurrent validity coefficients reported in the literature.

Discussion of Hypotheses Six's Findings: The relationship of the WISC-R verbal, performance and full scale IQ scores' concurrent validity with the automated RSPM followed the same relative pattern as the predictive relationship discussed in Hypothesis Five. It appeared, at least within the special, non-normal WISC-R automated RSPM group, which had been randomly selected from similar students who had been referred for psychoeducational testing due to academic difficulties assessed within this study, that verbal capabilities (as measured by the WISC-R verbal IQ scale) were more strongly related to the automated RSPM's IQ score than the WISC-R performance IQ scale.

The statistically significant relationship between the WISC-R verbal and full scale IQ's and the automated RSPM indicate that this newly developed instrument may be further developed into a useful screening device for predicting results of the WISC-R.

HYPOTHESIS SEVEN:

7. The results of the automated testing procedures will yield significant inter-relationships with the results of previously administered WISC-R subtest scores, or Verbal, Performance, and Full Scale scores when analyzed using the appropriate statistical procedures.

Statistical test employed: A correlation matrix was generated examining the group results of the WISC-R - automated RSPM group. The eight (8) RSPM variables and the WISC-R verbal, performance, full scale scores and age of subject were correlated. Please refer to the correlational matrix shown following the presentation of the multiple regression data.

A stepwise multiple regression procedure was employed on seven (7) different dependent variables. These were: Total Test Score, Total Test Time; Raven IQ; WISC-R Verbal IQ; WISC-R Performance IQ; WISC-R Full Scale IQ; and Age.

RSPM - TOTAL TEST SCORE

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	ARITH. SUBTEST	.44226	.44226	.0014

RSPM IQ SCORE

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	ARITH. SUBTEST	.45553	.45553	.0011

RSPM TOTAL TEST TIME

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	RESPON. TIME 2	.52671	.52671	.0003
2	ARITH. SUBTEST	.70305	.17634	.0055
3	RESPON TIME 1	.78387	.08082	.0264

WISC-R VERBAL IQ

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	FREED. FR. DIST.	.71368	.71368	.0000
2	SIMILAR. SUBT.	.87948	.16580	.0002
3	COMPRE. SUBT.	.93797	.05849	.0013
4	DIG. SPAN SUBT.	.96849	.02884	.0026
5	INFORM. SUBT.	.98110	.01429	.0058
6	VOCAB. SUBT.	.99648	.01537	.0000

WISC-R PERFORMANCE IQ

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	OBJ. ASS. SUBT.	.72594	.72594	.0000
2	CODING SUBTEST	.84313	.11718	.0024
3	PICT. COMP. SUB.	.91536	.07223	.0020
4	PICT. ARR. SUBT.	.96523	.04987	.0003
5	ARITHM. SUBTEST	.97580	.01057	.0258
6	SIMILAR. SUB.	.98278	.00698	.0389

WISC-R FULL SCALE IQ

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	FREED. FR. DIST.	.65329	.65329	.0000
2	OB. ASS. SUBT.	.84201	.18872	.0003
3	PICT. COMP. SUB.	.91964	.07763	.0012
4	COMPRE. SUBT.	.93870	.01905	.0474

AGE OF SUBJECTS

STEP	VARIABLE	R SQUARE	CHANGE R SQ	SIGNIFICANCE LEVEL OF F-TEST
1	RESPON. TIM.3	.33913	.33913	.0071
2	VOCABUL. SUBT.	.51502	.17589	.0238

HYPOTHESIS 7 - 1

	RTOTSC	RT1	RT2	RT3	ME1	ME2	RIQ
RSPM TOTAL SCORE (RTOTSC)	1.000	0.260	0.189	0.244	-0.079	-0.128	0.914
RESPONSE TIME INITIAL (RT1)	0.260	1.000	-0.135	0.045	-0.316	-0.178	0.262
RESPONSE TIME INTERMEDIATE (RT2)	0.189	-0.135	1.000	-0.064	-0.209	-0.218	0.220
RESPONSE TIME FINAL (RT3)	0.244	0.045	-0.064	1.000	0.382	0.824	0.282
MOTORIC EFFICIENCY ONE (ME1)	-0.079	-0.316	-0.209	0.382	1.000	0.498	-0.180
MOTORIC EFFICIENCY TWO (ME2)	-0.128	-0.178	-0.218	0.824	0.498	1.000	-0.037
RSPM DERIVED IQ (RIQ)	0.914	0.262	0.220	0.282	-0.180	-0.037	-0.270
AGE OF SUBJECT (AGE)	-0.024	-0.075	-0.194	-0.582	-0.028	-0.481	0.454
TOTAL TEST TIME (TSTIME)	0.432	0.309	0.726	0.133	-0.348	-0.155	0.599
WISC-R VERBAL IQ (VERBIQ)	0.499	-0.017	0.116	0.320	-0.204	0.276	0.462
WISC-R PERFORMANCE IQ (PERFIQ)	0.373	-0.010	0.132	0.710	0.092	0.544	0.618
WISC-R FULL SCALE IQ (FSIQ)	0.508	-0.008	-0.025	0.551	-0.079	0.455	0.618
FREEDOM FROM DISTRACT (XATTEN)	0.569	0.048	0.262	0.264	-0.409	0.046	0.659

HYPOTHESES 7 - 2

	AGE	TSTIME	VERBIQ	PERFIQ	FSIQ
RSPM TOTAL SCORE (RTOTSC)	-0.024	0.432	0.499	0.373	0.508
RESPONSE TIME INITIAL (RT1)	-0.075	0.309	-0.017	-0.010	-0.008
RESPONSE TIME INTERMEDIATE (RT2)	-0.194	0.726	0.116	0.116	0.132
RESPONSE TIME FINAL (RT3)	-0.582	0.133	0.320	0.710	0.551
MOTORIC EFFICIENCY ONE (ME1)	-0.028	-0.348	-0.204	0.092	-0.079
MOTORIC EFFICIENCY TWO (ME2)	-0.481	-0.155	0.276	0.544	0.455
RSPM DERIVED IQ (RIQ)	-0.270	0.454	0.599	0.462	0.618
AGE OF SUBJECT (AGE)	1.000	-0.357	-0.466	-0.533	-0.548
TOTAL TEST TIME (TSTIME)	-0.357	1.000	0.330	0.240	0.332
WISC-R VERBAL IQ (VERBIQ)	-0.466	0.330	1.000	0.548	0.903
WISC-R PERFORMANCE IQ (PERFIQ)	-0.533	0.240	0.548	1.000	0.850
WISC-R FULL SCALE IQ (FSIQ)	-0.548	0.332	0.903	0.850	1.000
FREEDOM FROM DISTRACT (XATTEN)	0.416	0.452	0.845	0.574	0.808

FINDINGS: The automated RSPM demonstrated a strong relationship with the Arithmetic subtest of the WISC-R. This was demonstrated in the comparison of all three of the RSPM dependent variables on which the stepwise multiple regression was performed. The automated RSPM parameters failed to demonstrate any significant relationships with WISC-R verbal, performance or full scale IQ scores when they were treated as dependent variables. Instead, the WISC-R subtests, which loaded on these scales, accounted for the significant variance determined by the multiple regression.

Discussion of Hypothesis Seven's Findings: The WISC-R's arithmetic subtest, a subtest which requires a subject to retain and process the greatest degree of verbal information of any of the subtests on the WISC-R, had a strong relationship with all of the RSPM parameters which were treated as dependent variables in the multiple regression. These relationships lend further credence to the aforementioned finding regarding the verbalization hypothesis which asserted that subjects verbally mediate their solutions to the RSPM.

HYPOTHESIS EIGHT:

8. The results of the automated testing procedures will yield significant inter-relationships with the previously derived WISC-R subtest factors derived through factor analysis.

Statistical test employed: A correlational matrix was generated using a rough equivalent of the Freedom from Distractibility factor and the CAA obtained parameters of the RSPM. The value of the Freedom from Distractibility factor was equal to the sum of the total Arithmetic and Digit Span subtests divided by 2 (Arith. + Dig. Span / 2). This value was used due to the limitations of the SPSS program's capabilities to perform more complex math computations. In addition, this value was determined to produce a metric equivalent to the other WISC-R scale scores. A critical value for the Pearson Product Moment Correlation was determined to be $r > .509$, thereby providing an alpha of .05 for the group size $N=20$. The following correlational matrix was generated:

HYPOTHESIS B - 1

	Freedom from Distractibility Factor
RSPM TOTAL SCORE (RTOTSC)	0.569
RESPONSE TIME INITIAL (RT1)	0.048
RESPONSE TIME INTERMEDIATE (RT2)	0.262
RESPONSE TIME FINAL (RT3)	0.264
MOTORIC EFFICIENCY ONE (ME1)	-0.409
MOTORIC EFFICIENCY TWO (ME2)	0.046
RSPM DERIVED IQ (RIQ)	0.659
AGE OF SUBJECT (AGE)	-0.416
TOTAL TEST TIME (TSTIME)	0.452
WISC-R VERBAL IQ (VERB IQ)	0.645
WISC-R PERFORMANCE IQ (PERF IQ)	0.574
WISC-R FULL SCALE IQ (FSIQ)	0.808
FREEDOM FROM DISTRACT (XATTEN)	1.000

FINDINGS: Statistically significant ($r > .65$) relationships were found between the RSPM IQ score ($r=.659$) and the Freedom from distractibility factor. In addition, the RSPM Total Score (which did not control for age of subjects) also yielded a significant ($r=.569$) relationship with the WISC-R freedom from distractibility factor. As would be expected the WISC-R verbal IQ ($r=.845$), full scale IQ ($r=.808$) and the Arithmetic ($r=.780$) and Digit Span ($r=.780$) subtests yielded higher correlations, as the Freedom from Distractibility Factor was directly drawn from these values. The vocabulary subtest's correlation ($r=.711$), which was also higher than the automated RSPM IQ score, appeared related to this subtests very strong loading on the WISC-R verbal IQ scale.

Discussion of Hypothesis Eight's Results: The automated RSPM demonstrated a significantly high loading on the Freedom from Distractibility factor. It is interesting to note that the RSPM IQ score yielded a higher correlation with this factor than the WISC-R's own performance IQ score ($r=.574$). In addition, the automated RSPM's

motoric efficiency one (1) parameter yielded a considerable negative correlation with the Freedom from Distractibility Factor (-.409). This finding suggests that the purposefulness in which subjects initially moved the automated RSPM's mouse in selecting their response appeared to have a relatively strong negative relationship with their ability to remain free from distractions when working on the WISC-R subtests demanding the most attention (Arithmetic and Digit Span).

The lack of further relationships between the other CAA obtained parameters and the WISC-R factor analytically derived Freedom from Distractibility Factor, may stem from the highly behaviorally specific nature of the RSPM measures relative to the more global nature of the two subtests which compose the WISC-R Freedom from Distractibility factor.

HYPOTHESIS NINE:

9. The unique psychometric parameters of subjects' three-phased item and total response times and two-phased measures of 'motoric efficiency' will provide statistically significant interrelationships when compared with other group and individual data obtained through automated and/or previously administered manually obtained psychometric data.

Statistical test employed: Due to the small sample size (N=20) of the automated RSPM - WISC-R group, it was determined that obtaining significant WISC-R interrelationships beyond those discussed above was very unlikely. Consequently a discriminate analysis was only run on the CAA data obtained from all subject's initial exposure to the CAA version of the RSPM (CAA - Manual, Manual - CAA, CAA - CAA; and WISC-R - automated RSPM). The subject's first exposure to the CAA format of the RSPM was the sole source of the discriminative analysis data for all cases.

The discriminate analysis was partitioned using the mean RSPM IQ score (IQ = 100.5) for all CAA subjects (N=85). Those subjects

with an IQ of 100 or below were placed in the low group and those with an IQ above 100 were placed in the high group of the discriminative analysis. Comparisons between the low and high scoring groups were made and significant ($\text{Alpha} = .05$) relationships were observed

DISCRIMINATIVE ANALYSIS

<u>PARAMETER</u>	<u>LOW GROUP</u>	<u>HIGH GROUP</u>	<u>TOTAL</u>
SAMPLE SIZE	43	42	85
<hr/>			
<u>MEAN SCORES OF:</u>			
TOTAL RSPM SCORE	31.7442	45.6667	38.6234
INITIAL RESPON TIME	10877.8605	12200.5952	11531.4470
INTERMED. RESPON. TIME	48209.5347	56935.9522	52521.4116
FINAL RESPONSE TIME	19640.3023	17082.4048	18376.3999
MOTORIC EFFICIENCY 1	72761.5586	77415.6670	75061.2354
MOTORIC EFFICIENCY 2	48469.1397	41303.8096	44928.6235
AGE OF SUBJECT	11.0258	11.2193	11.1214
TOTAL TEST TIME	26.5116	28.9048	27.6941
<hr/>			
<u>STANDARD DEVIATIONS OF:</u>			
TOTAL RSPM SCORE	6.9457	3.6202	8.9189
INITIAL RESPON. TIME	9951.0669	10221.4265	10047.3590
INTERMED RESPON TIME	19016.3938	17256.1072	18585.3435
FINAL RESPONSE TIME	6363.3077	6217.5137	6385.0867
MOTORIC EFFICIENCY 1	27037.3733	41758.2974	34958.6538
MOTORIC EFFICIENCY 2	19999.5459	16635.3645	18656.1162
AGE OF SUBJECT	1.2773	0.9504	1.1252
TOTAL TEST TIME	7.7502	6.1913	7.0846
<hr/>			

<u>VARIABLE</u>	<u>F SCORE</u>	<u>DEGREE OF SIGNIFICANCE</u>	
TOTAL RSPM SCORE	133.3440	**	< .01
INITIAL RESPONSE TIME	0.3655		-----
INTERMED RESPONSE TIME	4.9017	**	< .05
FINAL RESPONSE TIME	3.5118		< .10
MOTORIC EFFICIENCY 1	0.3738		-----
MOTORIC EFFICIENCY 2	3.2169		< .10
AGE OF SUBJECT	0.6250		-----
TOTAL TEST TIME	2.4667		-----

FINDINGS: Two significant relationships were observed ($\alpha = .05$).

The Total RSPM score demonstrated a very significant relationship with the IQ score, as would be expected. In addition, the CAA obtained measure of Intermediate Response Time, the time a subject takes from initially moving the cursor until he clicks on his initial choice, was also found to significantly differentiate those subjects who scored on either half of the mean. In addition, two other CAA obtained measures demonstrated near significant levels of discriminability at the pre-determined alpha level of .05. Actually these scores demonstrated a significant relationship at an alpha level of .10, however, the a priori determination to use an alpha level of .05 precludes labeling them as significant findings. These almost

significant results were Final Response Time and Motoric Efficiency Two

Discussion of Hypothesis Nine's Results: The determination of significant results on one CAA obtained measure, and near significant results on two other measures indicated that the CAA format of evaluation not only yielded comparable results to the manual versions of the same test, but furthermore yielded highly discriminative psychometric parameters unobtainable without computer mediated techniques.

Examination of the directionality of the three significant/near significant findings revealed much about the behavioral determinants which differentiated the more successful subjects from their lower scoring counterparts. The above average subjects spent more time in initially choosing and clicking on their initial response item than did the below average subjects. This was revealed by the higher average Intermediate Response Time for the above average group. This same group more rapidly determined that their initially chosen responses were correct and thereby more rapidly locked in their responses. This

phenomenon was demonstrated by the lower average Final Response Time for the above average group. In addition, this group demonstrated a much higher degree of purposefulness in moving their cursor from their initial choice to the area designated for locking in responses. This finding was demonstrated by the lower average Motoric Efficiency Two scores for the above average group of subjects.

Thus these significant/near significant computer obtained cumulative measures of item by item problem-solving behavior affirmed the intuitive assumption that the more successful subjects worked more carefully prior to making a choice and then demonstrated a higher degree of confidence and efficiency once their choice has been made. These findings suggest that further development of measures like the three component response time and measures of motoric efficiency may prove to be very significant psychometric parameters in CAA devices in the future.

CHAPTER VI

SUMMARY, CONCLUSION, AND IMPLICATIONS FOR FURTHER RESEARCH.

A. SUMMARY

The study considered the feasibility, practicality and degree of concurrence of Computer Assisted Assessment (CAA) with that of traditional paper and pencil administered (manual) intelligence tests. The study investigated the test-retest reliabilities of the newly developed CAA instrument, the interrelationships between the CAA instrument and the manual (paper and pencil) version of the same test (Raven Standard Progressive Matrices - RSPM), and the capability of the CAA test to predict the scores of a different intelligence test, the Wechsler Intelligence Scale for Children - Revised (WISC-R)

In addition, the study developed three measures of response time which were designed to analyze students' personal styles of cognitive problem solving by fragmenting their responses into the fundamental behavioral aspects required for the completion of each test item and timing these responses in an extremely precise, covert manner. An additional means of assessing the students' purposefulness in performing the fragmented essential elements of each test item was also developed, using the newly developed capabilities of the Macintosh computer. This measure of 'motoric efficiency', which measured the linear movement the student displayed while completing the CAA test, was postulated to provide significant relationships with the relative success students demonstrated in performing the CAA version of the RSPM.

The study consisted of two phases. The first investigated the psychometric interrelationships between the newly developed CAA test and the manual version of the same test, the RSPM. The second portion of the study examined the interrelationships between the CAA instrument and the WISC-R. In addition, data from the computer

administration of the RSPM during both portions of the study was pooled and examined separately to investigate the psychometric properties of the CAA derived measures of response times and motoric efficiency scores.

Nine hypotheses were postulated of which eight were tested. Significant results were found within all eight of the tested hypotheses.

Within the first phase of the study, which examined the psychometric properties of the CAA version with itself and with the manual version, numerous significant and near significant findings were obtained. The CAA version of the RSPM was found to be a reliable instrument when examined in a test-retest format with a two-week interval between test sessions. A test-retest correlation of $r = .876$ was found in the two week test-retest reliability study of the CAA version of the RSPM (CAA then CAA group). This finding was quite similar to the test-retest reliabilities reported for the manual version of the RSPM which range between $r = .55$ to $.93$ for short-term reliability (Raven, 1983, p. SPM7). This study's obtained reliability coefficient actually falls at the upper range of reported manual RSPM

test-retest reliabilities. This high degree of test-retest reliability may be, in part, the result of examiner effects being eliminated by the CAA format of testing, which provides a consistent degree of feedback with each administration.

The CAA version of the RSPM when compared with the manual version in a counter-balanced, test-retest experimental design (CAA then manual versus manual then CAA) demonstrated a comparable degree of practice effect as demonstrated by post-test scores. No significant order effects were found which affected the respective total scores obtained by either order of test presentation.

This counter-balanced design format when statistically analyzed by using a Multiple Analysis of Variance Test (MANOVA) did demonstrate a significant order effect between the CAA first group and the paper and pencil first group on the amount of total time subjects spent on their pretest. The CAA first group spent a significantly greater amount of time on the pretest portion of the counter-balanced design than did the manual group. The CAA format may have required more time for subjects to complete the test because of the introduction of the two component response format,

wherein the subjects had to initially choose their answer for each item and then 'lock in' their answer. In addition, the hardware/software interaction of the test introduced an approximate one-second pause between each item's presentation. A precise determination of the factors which attributed for the significant difference between the pretest times of the CAA and the manual groups cannot be determined by this study due to the lack of a manual then manual group. Omission of this group was made at this study's inception, due to the strong research base which clearly demonstrated the reliability of the manual version of the RSPM. Unfortunately, it was not initially recognized that inclusion of this group would have yielded other information of relevance to the study.

A final analysis of the psychometric properties of the newly developed CAA RSPM was performed using a discriminate analysis. The subjects whose scores were grouped for the discriminate analysis came from the aforementioned groups which were: CAA then manual; manual then CAA, and CAA then CAA. In addition the WISC-R then CAA RSPM group's CAA scores were also included. The scores obtained on all of these subjects' first administration of the

CAA RSPM were grouped (N=85) and a mean score of 100.5 was obtained. The discriminate analysis was partitioned by this mean score, wherein subjects with an IQ of 100 or below were placed in the low group and those with an IQ above 100 were placed in the high group of the discriminate analysis. Comparisons between the low and high scoring groups were made and two significant ($\text{Alpha} = .05$) and two other near significant ($\text{Alpha} = .10$) psychometric determinates for high or low group placement were observed.

As would be expected, the Total CAA RSPM score significantly differentiated between high and low scoring subjects. In addition, one of the CAA derived response time measures, intermediate response time, also displayed a significantly different value for the high scoring students versus the low scoring students. Intermediate response time measured the amount of time that subjects' required from initially moving the mouse-controlled, screen cursor (arrow) towards a response choice until they actually indicated their choice by positioning the cursor on an answer and clicking on a button on the computer's mouse (a device used to provide human/machine interaction). The discriminate analysis indicated that high scoring

subjects spent significantly more time on this portion of the three components of response time (initial, intermediate and final) than did low scoring subjects. This finding suggests that high scoring students worked more carefully and took more time in selecting their answers than did low scoring students.

Some other near significant ($p < .10$) findings provide further interesting data regarding the quantitative CAA derived measures of the subjects' style of working on the CAA RSPM. Near significant differences were found between high and low scoring students in final response time and the second phase of motoric efficiency (motoric efficiency 2). The high scoring students were found to spend less time than the low scoring students in the final response time phase which involved 'locking in' their responses after selecting their choice for each item. In addition, the higher scoring students demonstrated a lower, more purposeful rate of motor output when moving their mouse-mediated cursor from their, initial choice for each item until 'locking in' their responses. In sum, the higher scoring students were found to spend more time initially choosing their answers (intermediate response time) than lower scoring

students. However, once the higher scoring group of students had made their determination of the correct answers, they more rapidly, (final response time) and with a greater degree of motoric purposefulness (motoric efficiency 2) completed that item's response and moved on to the next item. Thus these three CAA derived psychometric measures indicate what is intuitively obvious, yet which heretofore has only been qualitatively observed.

The second phase of the study investigated the psychometric interrelationships between the WISC-R and the CAA RSPM. Four hypotheses had been postulated regarding the WISC-R - CAA RSPM interrelationships. The first asserted that the CAA RSPM would significantly predict, in a post hoc manner, subject's previously administered WISC-R scores. A multiple regression analysis was performed treating the WISC-R IQ scores as dependent variables. The CAA obtained measures (response times and motoric efficiency) and the RSPM derived IQs (interpolated from the percentile scores of the manual RSPM) were then examined to determine if any accounted for a significant amount of the respective WISC-R scores' variance.

Significant interrelationships were found between the RSPM IQs and the WISC-R Verbal, Performance and Full Scale IQ scores. In fact, the RSPM IQs demonstrated an even higher loading on the WISC-R Verbal and Full Scale IQ scores than on the Performance IQ scores. This finding violated the intuitive consideration of the RSPM as a 'non-verbal test'. This unexpected finding suggested that verbalizing the solution to the CAA RSPM items may play a large role in cognitive strategies used in performing the RSPM. This so-called "verbalization hypothesis" has previously been discussed in the literature regarding the manual RSPM (Deutsch, Katz and Jensen, 1968, Burke and Bingham, 1968, and Raven, 1983). In addition, the CAA derived measure of final response time was found to be very significantly related to the WISC-R Performance IQ scores. This relationship accounted for 58% of the variance of the WISC-R performance IQ scale, which was a stronger relationship than the CAA RSPM IQ score yielded. This finding suggests that the CAA obtained measure of final response time, which previously was discussed for its nearly significant role in discriminating among high and low scorers on the

CAA RSPM, also relates very strongly ($p < .0002$) with the WISC-R performance IQ score.

The CAA obtained parameters of motoric efficiency 1 and motoric efficiency 2 were found to significantly account for the variance of the WISC-R full scale IQs and verbal IQs respectively. These findings, coupled with the very significant relationship of final response time with the WISC-R performance IQ score, suggest that CAA based testing may eventually yield predictions of academic success (IQ's) comparable with those of highly regarded manual IQ tests like the WISC-R. This would be possible due to CAA devices ability to measure parameters like response times and motoric efficiency which may account for additional variance within the 'intelligence factor' which had been previously untapped by conventional intelligence tests.

The second hypothesis examined in the second phase of the study examined the correlation between the WISC-R and the CAA RSPM in an effort to access the concurrent validity between the two instruments. The intercorrelations between the RSPM's interpolated IQ scores and the WISC-R scores were. Verbal = .599;

Performance = 462 and Full Scale IQ = 618. These interrelationships displayed the same relationship as the multiple regression findings, in that a stronger relationship was found between the WISC-R verbal and full scale scores than the performance score.

The WISC-R with the CAA RSPM comparisons also discovered a significant relationship between the WISC-R arithmetic subtest and the RSPM IQ scores and RSPM total time on test, when examined using a stepwise multiple regression analysis. These findings suggest that the underlying cognitive processes used by subjects in performing the CAA RSPM may be similar to those required to complete the WISC-R arithmetic subtest.

A final comparison of the WISC-R with the CAA RSPM was performed examining the relationship between the previously factor analysis derived WISC-R freedom from distractibility factor and the CAA RSPM interpolated IQ scores. A correlation of .65 was found between the RSPM IQ scores and this study's defined value of freedom from distractibility (WISC-R arithmetic subtest scaled scores + digit span subtest scaled scores divided by 2). This correlation was found to exceed all WISC-R subtest scaled score correlations with the

exception of arithmetic and digit span (the two subtests which make up this factor) and vocabulary. Therefore, the RSPM IQ relates very strongly with factors which loaded on the freedom from distractibility factor which had been previously derived on the WISC-R. This finding suggests that subjects' attention to task as demonstrated on the WISC-R freedom from distractibility factor also relates strongly with subjects' RSPM IQ scores.

B. CONCLUSION

The study attempted to demonstrate the feasibility, practicality, and degree of concurrence of Computer Assisted Assessment (CAA) with that of traditional paper and pencil administered (manual) Intelligence tests. Examination of the reliability of the CAA instrument indicated that it demonstrated a high degree of test-retest reliability ($r = .876$) when a two week period was used between test sessions.

A determination of the concurrence between the CAA and manual version of the RSPM indicated that the two different test formats yielded extremely similar results. No significant order effects between scores was detected in a counter-balanced test-retest design. A significant order effect was found between the two different test formats in the total time of test. This finding indicated that the CAA version of the test required significantly more time to complete than the manual version of the test.

The concurrence between the CAA RSPM and the WISC-R, when compared in a post hoc manner, was also found to be relatively high. This is similar to the findings for the manual version of the RSPM

reported in the research literature regarding these test instruments. Thus, the CAA RSPM appeared to be a relatively good predictor of the WISC-R, particularly in regards to the verbal and full scale IQ scores of the WISC-R.

Overall, the study demonstrated the very high test-retest reliability of the CAA version of the RSPM, the strong concurrence with the manual version of the same test, and an adequate degree of predictability of the WISC-R when examined in a post hoc manner.

The study also examined the relationships of the newly developed CAA derived measures of response time and motoric efficiency. Numerous interrelationships between these measures of subjects' actual cognitive behavioral styles and the traditional psychometric scores of the RSPM and the WISC-R were found. These findings suggested that the efficacy of CAA derived measures, which have heretofore been unattainable by practicing school psychologists, may eventually allow for increased accuracy in test interpretation and in predicting students potential for academic success.

This study demonstrated that a computer based, automated testing device could be created from an 'off-the-shelf' computer, the

Apple Macintosh. Results comparable to manual versions of the test are also possible to obtain. In addition, unique and meaningful psychometric parameters which may aid in accounting for a greater degree of the variance in predicting the probability of academic success may be possible through the use of CAA obtained measures like response times and motoric efficiency

In conclusion, Computer Assisted Assessment (CAA) has been demonstrated to yield results that are comparable to manually administered intelligence tests. The additional psychometric parameters derived by the CAA device like motoric efficiency and response time may allow for eventual development of CAA devices that yield superior predictions of intellectual ability than the current manual instruments

C. IMPLICATIONS FOR FUTURE RESEARCH

Future development of CAA testing devices will probably consist of two different phases of research. The first phase will be similar to the current study's form of investigation, wherein existing manual tests are automated and compared with the results of the manual test. More research of this nature is required to ascertain any unique psychometric effects produced by the computer/human interaction.

An important area of exploration within this first phase of research will be a determination of the effects produced by perceptual deficits. Integration and automation of existing tests like the Motor-Free Visual Perception Test (Colarusso and Hammill, 1972) and the current automated version of the RSPM may offer a ready means to discern the unknown relationship between perceptual factors and CAA presentation format of psychometric stimuli.

Following further determination of the unique effects produced by CAA testing, new tests specifically developed for the CAA format of presentation will be developed. This area of inquiry appears most

promising due to the ability of the CAA format to assess human responses in ways which have heretofore been impossible. Consequently, newly developed CAA devices which employ previous forms of assessment coupled with new CAA derived measures may eventually lead to instruments which can more accurately predict human behavior than existing manual tests. The integration of behaviorism, cognitive psychology, and psychometrics, all necessary elements to the development of these putative CAA devices, may well be brought about by the development of these CAA devices.

Within the area of school psychology specifically, future researchers may wish to explore refinement of psychometric assessment wherein diagnosis, prescription, and remediation of children's difficulties in learning are all part of a continuum which is monitored and mediated by specially designed computer systems. The advent of inexpensive, laser-read storage systems for electronic media which can hold several hundred Megabytes of information is approaching within months. Consequently, entire curriculums of remedial materials for specific categories of educational disability could be created and stored on such disks. Thus, a child's

learning difficulties could be precisely determined, a teacher/computer based remediation program could be prescribed and the computer could facilitate the teacher trained to assist children with such educational difficulties

The productivity of teachers and students could thus be enhanced through the use of computer based diagnosis, prescription and remediation

The productivity of psychologists' may also be dramatically enhanced by CAA devices through reducing the time spent in rote testing situations. Such devices should be viewed and marketed as a supplement to psychological services and never as a replacement for such services. The use of such devices must always be under the control and supervision of a qualified examiner. It is especially important that test interpretation of CAA obtained results be performed by qualified personnel with appropriate graduate level training and expertise in the psychological areas being measured

The future research and professional implications for CAA devices are considerable. The capabilities and promise of such

devices appear limited only by the commitment, imagination and resources applied to their use and further development by psychologists.

A - 1

APPENDIX ARSPM DERIVED IQ TABLE

ACTUAL AGE RANGE(MONTHS)	76-81	82-87	88-93	94-99
AGE GROUPING	6.5	7.0	7.5	8.0
NUMBER CORRECT				
8	BELOW 75			
9	75	BELOW 75		
10	80	75	BELOW 75	BELOW 75
11	84	78	75	75
12	87	80	80	77
13	90	85	84	79
14	94	90	87	80
15	97	94	90	84
16	100	97	92	87
17	101	100	93	90
18	102	101	95	92
19	104	102	96	93
20	106	104	98	94
21	108	106	99	95
22	110	108	100	97
23	111	110	101	98
24	112	111	102	99
25	113	112	103	100
26	115	113	105	101
27	116	114	106	102
28	117	116	107	103
29	118	117	108	105
30	120	118	110	106
31	121	119	112	107
32	123	120	114	108
33	125	122	116	110
34	ABOVE 125	125	118	112
35		ABOVE 125	120	114
36			122	116
37			125	118
38			ABOVE 125	120
39				122
40				125
41				ABOVE 125
42				
43				
44				
45				

A - 2

ACTUAL AGE RANGE (MONTHS)	100-105	106-111	112-117	118-123
AGE GROUPING	8-5	9-0	9-5	10-0
NUMBER CORRECT				
9				
10	BELOW 75			
11	75			
12	76			
13	77	BELOW 75	BELOW 75	
14	78	75	75	
15	79	76	76	
16	79	78	77	BELOW 75
17	80	80	78	75
18	82	82	79	76
19	84	83	80	77
20	86	84	81	78
21	86	85	82	78
22	90	87	83	79
23	91	88	84	80
24	92	89	86	81
25	93	90	87	82
26	94	92	88	83
27	96	93	89	84
28	97	94	90	86
29	98	95	92	87
30	99	97	93	88
31	100	98	94	89
32	102	99	95	90
33	104	100	97	92
34	106	102	98	94
35	108	104	99	96
36	110	106	100	98
37	112	108	102	99
38	115	110	104	100
39	117	112	106	102
40	120	115	108	105
41	122	117	110	107
42	125	120	113	110
43	ABOVE 125	122	117	112
44		125	120	115
45		ABOVE 125	122	117
46			125	120
47			ABOVE 125	122
48				125
49				ABOVE 125
50				
51				

A - 3

ACTUAL AGE RANGE (MONTHS)	124-125	130-135	136-141	142-147
AGE GROUPING	10-5	11	11-5	17
NUMBER CORRECT				
21	BELOW 75			
22	75			
23	76	BELOW 75		
24	77	75	BELOW 75	
25	78	76	75	BELOW 75
26	79	77	76	75
27	80	78	77	76
28	82	79	78	77
29	84	80	78	78
30	86	82	79	79
31	88	84	80	80
32	89	86	82	82
33	90	88	84	84
34	92	90	86	86
35	94	92	88	88
36	96	94	90	89
37	98	96	92	90
38	99	98	94	93
39	100	99	96	95
40	102	100	98	98
41	105	102	100	100
42	107	105	102	102
43	110	107	105	104
44	112	110	107	106
45	115	112	110	108
46	117	115	112	110
47	120	117	115	112
48	122	120	117	115
49	125	122	120	117
50	ABOVE 125	125	122	120
51		ABOVE 125	125	122
52			ABOVE 125	125
53				ABOVE 125
54				
55				

A - 4

ACTUAL AGE RANGE/MONTHS	140-153	154-159	160-165	166-171
AGE GROUPING	12-5	13	13-5	14
NUMBER CORRECT				
26	BELOW 75			
27	75	BELOW 75		
28	76	75	BELOW 75	
29	77	76	75	BELOW 75
30	78	77	76	75
31	79	78	77	76
32	80	79	78	77
33	81	80	79	78
34	84	82	79	76
35	86	84	80	79
36	88	86	82	80
37	89	88	84	82
38	90	89	86	84
39	93	90	89	86
40	95	92	89	88
41	96	95	90	89
42	100	98	94	90
43	102	100	97	94
44	104	101	100	97
45	106	102	102	100
46	108	104	104	102
47	110	106	106	104
48	111	108	108	106
49	115	110	110	108
50	117	113	112	110
51	120	117	115	112
52	122	120	117	115
53	125	122	120	117
54	ABOVE 125	125	125	120
55		ABOVE 125	ABOVE 125	125
56				ABOVE 125
57				
58				
59				
60				

A - 5

ACTUAL AGE RANGE (MONTHS)	172-177	178-183	184-189
AGE GROUPING	14.5	15.0	15.5
NUMBER CORRECT			
26			
27			
28			
29			
30			
31			
32	BELOW 75	BELOW 75	BELOW 75
33	75	75	75
34	77	77	77
35	79	79	79
36	80	80	80
37	82	82	82
38	84	84	84
39	86	86	86
40	88	88	88
41	89	89	89
42	90	90	90
43	93	92	92
44	95	94	94
45	98	96	96
46	100	98	98
47	102	100	100
48	105	102	102
49	107	105	105
50	110	107	107
51	112	110	110
52	115	112	112
53	117	115	115
54	120	117	117
55	122	120	120
56	125	122	122
57	ABOVE 125	125	125
58		ABOVE 125	ABOVE 125
59			
60			

APPENDIX BSCHOOL DIVISION RESEARCH FORMS

VIRGINIA BEACH CITY PUBLIC SCHOOLS

SCHOOL ADMINISTRATION DEPARTMENT • 1000 COLLETT DRIVE • VIRGINIA BEACH, VIRGINIA 22192

TELEPHONE (703) 681-1111
FAX (703) 681-1111

Name of Child

Dear Parent:

Your child has been randomly chosen to participate in an experiment evaluating a new computer administered test. With your permission, your child will be given the opportunity to work with an interactive computer-based test measuring intellectual potential. In addition, a pencil and paper version of a similar test will be administered. The total amount of time for the two test sessions will be approximately one hour.

The tests will be administered by myself, a certified school psychologist employed at your child's school. The results of the study will be kept in complete confidence. None of your child's test scores will be reported to school personnel due to the experimental nature of the test.

This study is a part of my dissertation, which will partially fulfill the requirements for my doctoral degree in school psychology and counseling at the College of William and Mary. The results of this preliminary study may ultimately enhance the efficiency and accuracy of Virginia Beach psychologists in assessing the intellectual potential and development of children.

I greatly appreciate your consideration of this letter and your cooperation in this research. Please indicate your approval/disapproval of your child's participation in this study on the enclosed permission form and return it to the school your child attends.

Sincerely,

Michael A. Buxton

Michael A. Buxton
Certified School Psychologist

d1h

Enclosure



VIRGINIA BEACH CITY PUBLIC SCHOOLS

SCHOOL ADMINISTRATION BUILDING • P. O. BOX 6035 • VIRGINIA BEACH, VIRGINIA 23456

F. E. BRICKILL
SUPERINTENDENT OF SCHOOLS

Name of Child

Dear Parent:

Your child has been chosen at random from children who have been tested during the 1984-85 school year. With your permission, your child will be given the opportunity to work with a new computer administered test of intellectual potential. The total testing time will be approximately one-half hour. The testing will have no effect on your child's eligibility for Special Education.

The computer-based test will be administered by a certified Virginia Beach school psychologist. The computer-derived test results of your child will not be reported to school personnel due to the experimental nature of the test. With your permission, these results will be compared with the test results of your child, which were administered earlier this year. These results will be kept in complete confidence by myself and will be reported only as a group result.

This study is a part of my dissertation, which will partially fulfill the requirements for my doctoral degree in school psychology and counseling at the College of William and Mary. The results of this preliminary study may ultimately enhance the efficiency and accuracy of Virginia Beach psychologists in assessing the intellectual potential and development of children.

I greatly appreciate your consideration of this letter and your cooperation in this research. Please indicate your approval/disapproval of your child's participation in this study on the enclosed permission form and return it to the school your child attends.

Sincerely,



Michael J. Buxton
Certified School Psychologist

dih

Enclosure



I GIVE PERMISSION for my child _____
to participate in the computer administered testing study of Michael J.
Buxton described in the letter I received from him. I understand that
due to the experimental nature of the test no results of my child's
performance will be released to school personnel.

Date

Signature of Parent/Guardian

Telephone Number

I DO NOT GIVE PERMISSION for my child _____
to participate in the computer administered testing study of Michael J.
Buxton as described in the letter I received from him.

Date

Signature of Parent/Guardian

Telephone Number

VIRGINIA BEACH CITY PUBLIC SCHOOLS

SCHOOL ADMINISTRATION BUILDING • P. O. BOX 6038 • VIRGINIA BEACH, VIRGINIA 23456

F. E. BRICKELL
SUPERINTENDENT OF SCHOOLS

Dear Principal:

Your school has been chosen for a research study on the effectiveness of fully automated, computer-assisted, intellectual testing. In the interest of possibly enhancing the productivity of school psychologists, a computer-assisted assessment test has been developed in conjunction with the preparation of my dissertation, which will partially fulfill the requirements for my doctoral degree in school psychology and counseling at the College of William and Mary.

The psychologist assigned to your school has offered to participate in this study, with your permission. The computer-based test is fully automated and self-scoring. It will require no more than 15 minutes of the psychologist's time to set up, brief the child, and disassemble the equipment. An IQ score with high theoretical concurrence with WISC-R scores will be derived. This score will be obtained by having the psychologist explain the "pointing response" of the test and allowing the child to work with the machine individually for approximately 30 minutes. No scoring or reporting time will be required of the psychologist. I am sure you recognize how such a computerized instrument could enhance your psychologist's delivery of services to your school.

With your permission, one to three children in your school who have already been given the WISC-R during the 1984-85 school year will be randomly chosen. If you agree to participate in this study, the parents of these children will be given the enclosed permission letter by your school psychologist. Upon receipt of the signed permission form from the parents, the children will be administered the automated test. This process will take approximately one-half hour.

The test results will not be released to school personnel or to parents due to the experimental nature of the automated test. Hopefully, validation studies like this one will lead to a proven automated device which will assist psychologists in screening intellectual and academic areas of children.

Enclosed is the permission to test letter which your school psychologist will send to one to three children within your school. Also enclosed is a copy of the authorization form from the Office of Planning, Assessment, and Resource Development with the Virginia Beach City Schools and a research summary of my proposed study.

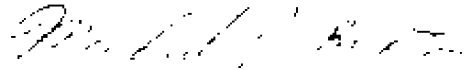


Letter to Principals
Page 2

If you have any questions or reservations about this study, please call me at 427-4856 during school hours or at my home (464-1051) in the evening. I have designed this study to require minimal usage of time involving students and school personnel and yet provide very useful educational information.

Your cooperation with this study is greatly appreciated. Please indicate your approval/disapproval of your school's participation in this research on the enclosed permission form and return it to me.

Sincerely,



Michael J. Buxton
Certified School Psychologist

dlh

Enclosures

I GIVE PERMISSION for my school _____
to participate in the computer administered testing study of Michael J.
Buxton described in the letter I received from him. I understand that,
due to the experimental nature of the test, no individual results of
the testing will be released to school personnel.

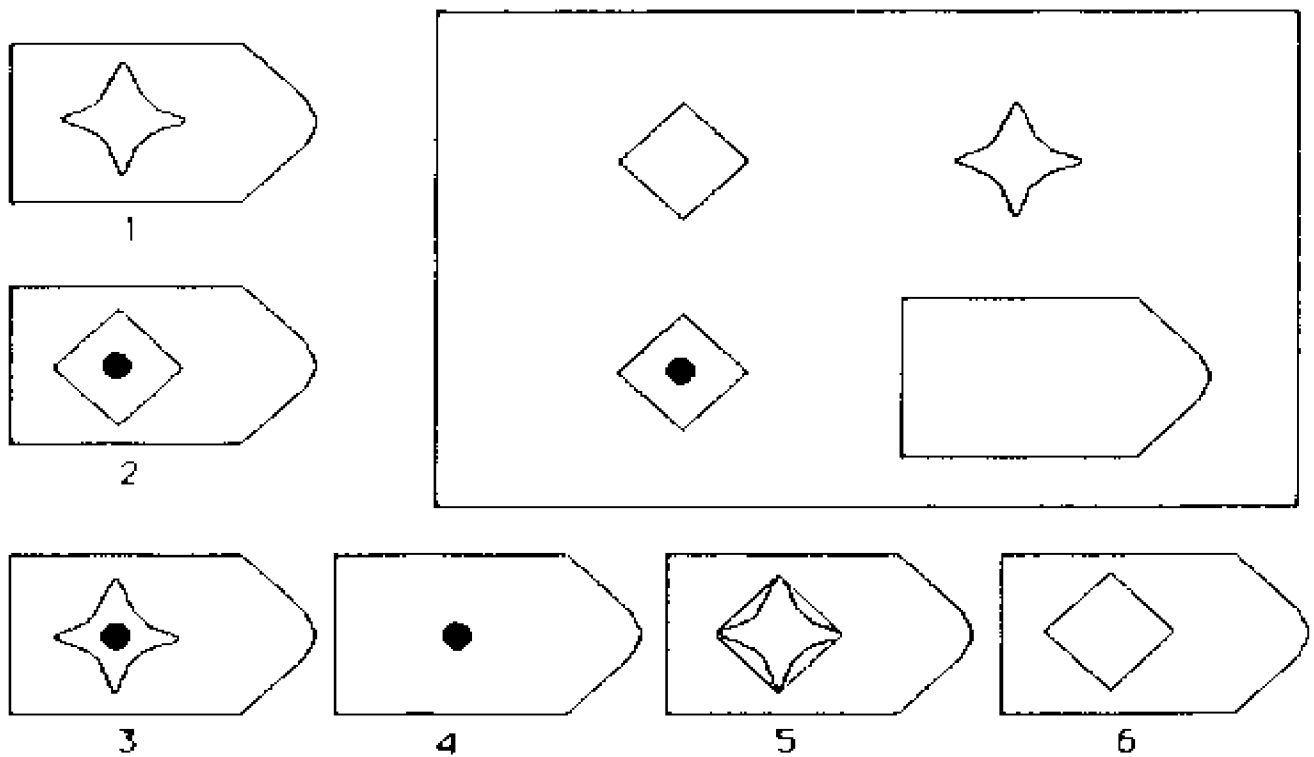
Date

Signature of School Principal

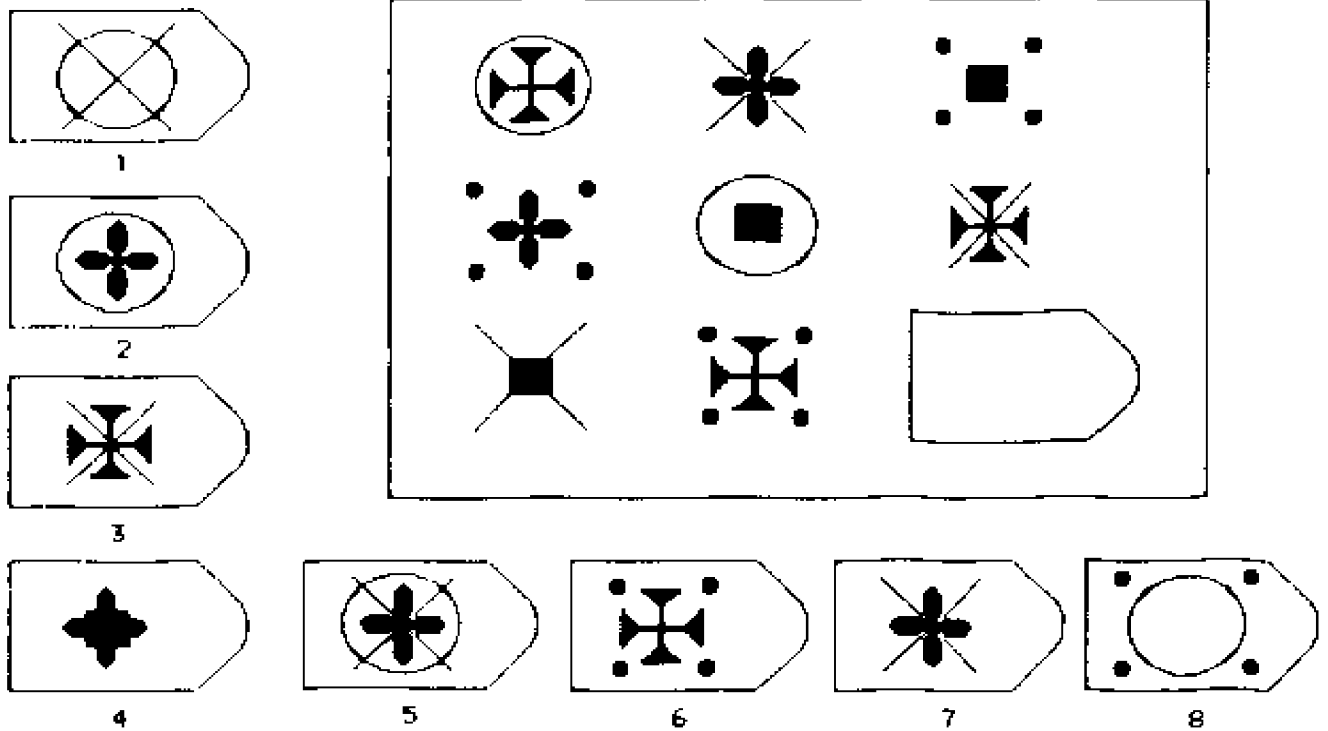
I DO NOT GIVE PERMISSION for my school _____
to participate in the computer administered testing study of Michael J.
Buxton described in the letter I received from him.

Date

Signature of School Principal

APPENDIX C**EXAMPLE OF MACPAINT CAA RSPM IMAGES
ITEM B - 10 (6 RESPONSES)**

EXAMPLE OF MACPAINT CAA RSPM IMAGE
ITEM D - 10 (8 RESPONSES)



REFERENCES

- Acker, W. In support of microcomputer based automated testing - a description of the Maudsley Automated Psychological Screening Tests (MAPS) British Journal on Alcohol and Alcoholism, 1980a, 15, 144-147
- Acker, W. A microcomputer administered neuropsychological assessment system for use with chronic alcoholics. Substance and Alcohol Actions/Misuse , 1980b, 1, 545-550.
- Acker, W. A computerized approach to psychological screening - The Bexley - Maudsley Automated Psychological Screening and The Bexley - Maudsley Category Sorting Test International Journal of Man-Machine Studies, 1983, 18, 361-369
- Anastasi, A. Psychological Testing - Fifth Edition. New York: Macmillan Publishing Co., Inc., 1982
- Apple Computer Co. Inside Macintosh, Apple Computer Co., Cupertino, California. Author, 1985.

Beaumont, J. G. Automated administration of standardized psychological tests by APPLE Applications of APPLE Symposium, British Psychological Society, London Conference, 1980

Beaumont, J. G. System requirements for interactive testing
International Journal of Man-Machine Studies, 1982, **17**,
311-320.

Brierley, H. A fully automated intellectual test. British Journal of Social and Clinical Psychology, 1971, **10**, 286-288.

Brown, A. L. & Camplone, J. C. Inducing flexible thinking: the problem of access. In M. Friedman & J. P. Das & N. O'Connor (Eds.), Intelligence and Learning New York: Plenum Press, 1981.

Burke, H. R. & Bingham, W. C. Raven's Progressive Matrices: More on construct validity. Journal of Psychology, 1969, **72**, 247-251.

Calvert, E. J. and Waterfall, R. C. A comparison of conventional and automated administration of Raven's Standard Progressive Matrices. International Journal of Man-Machine Studies, 1982, **17**,
305-310

- Campbell, D. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin Co., 1963.
- Colarusso, R. P. and Hammill, D. D. Motor-Free Visual Perception Test Manual. Novato, California: Academic Therapy Publications., 1972.
- Colles, J. M. Automated standard head injury tests. M.Sc Dissertation, University of Manchester Institute of Science and Technology, 1975.
- Denner, S. J. Automated psychological testing: a review. British Journal of Social and Clinical Psychologists, 1977, 16, 175-179.
- Elithorn, A., Cooper, R. & Lennox, R. Assessment of psychotropic drug effects. In J. Crooks & I. H. Stevenson (Eds.), Drugs and the Elderly. London: Macmillan, 1979.
- Elithorn, A., Kerr, M. & Jones, D. A binary perceptual maze. American Journal of Psychology, 1963, 76, 506-508.

- Elithorn, A., Mornington, S. and Stavrou, A. Automated psychological testing: some principles and practice. International Journal of Man-Machine Studies, 1982, 17, 247-263.
- Elithorn, A., Powell, J., Telford, A. & Cooper, R. An Intelligent terminal for Automated Psychological Testing and Remedial Practice. In R. L. Grimsdale & H. C. A. Hankins (Eds.), Human Factors and Interactive Displays. Buckingham: Network, 1980.
- Elwood, D. L. Automation of psychological testing. American Psychologist, 1969, 24, 287-289.
- Elwood, D. L. Test retest reliability and cost analyses of automated and face to face intelligence testing. International Journal of Man-Machine Studies, 1972a, 4, 1-22.
- Elwood, D. L. Automated WAIS testing correlated with face to face Wais testing: a validity study. International Journal of Man-Machine Studies, 1972b, 4, 129-137.
- Evans, C. R. Chatting with computers. Electronics and Power, 1975, 21, (7), 430-432.

- Exton-Smith, A. N. A psychometric test battery for evaluating the action of psychotropic drugs. In G. Barbagallo & A. N. Exton-Smith (Eds.), The Aging Brain: Neurological and Mental Disturbances. New York: Plenum Press, 1980.
- Flowers, K. An automated short-term visual memory test. Bulletin of the British Psychological Society, 1968, **21**, 102.
- Gedye, J. L. Experience with an Automated Psychological Testing system designed for clinical use. Bulletin of the British Psychological Society, 1966, **19**, 18.
- Gedye, J. L. A teaching machine programme for use as a test of learning ability. In D. Uwin & J. Leedham (Eds.), Aspects of Educational Technology, London: Methuen, 1967.
- Gedye, J. L. The development of a general purpose psychological testing system. Bulletin of the British Psychological Society, 1966, **21**, 101-102.
- Gedye, J. L. & Gaines, B. R. Medical aspects of programmable audio-visual displays. International Conference on Medico-Biological Engineering, Stockholm, p.276, 1967.

- Gedye, J. L. & Miller, E. The automation of psychological assessment. International Journal of Man-Machine Studies, 1969, 1, 237-262.
- Gedye, J. L. & Miller, E. Developments in automated testing systems. In P. Mittler (Ed.), The Psychological Assessment of mental and Physical Handicaps. London: Methuen, 1970.
- Gedye, J. L. & Wedgewood, J. Experience in the use of a teaching machine for the assessment of senile mental changes. Proceedings of the 7th International Congress of Gerontology, Vienna, p 84, 1966.
- Gedye, J. L., Exton-Smith, A. N. & Wedgewood, J. A method of measuring mental performance in the elderly and its use in a pilot clinical trial of meclofenoxate in organic dementia (preliminary communication). 1972, Age and Ageing, 1, 74-80.
- Grant, D. A. & Berg, E. A. A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. Journal of Experimental Psychology, 1948, 38, 408-411.
- Hunt, E. Mechanics of verbal ability. Psychological Review, 1978, 85, 109-130.

Hunt, E. Hunt Intelligence as an information-processing concept. British Journal of Psychology, 1980, 71, 449-474.

Kaufman, A. S. Factor analysis of the WISC-R at eleven age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 1975, 43, 135-147.

Kaufman, A. S. Intelligent testing with the WISC-R. New York: John Wiley & Sons, 1979.

Knights, R. M., Richardson, D. H. & McNarry, L. R. Automated vs. clinical administration of the Peabody Picture Vocabulary Test and Coloured Progressive Matrices. American Journal of Mental Deficiency, 1973, 78(2), 223-225.

Levy, R. & Post, F. The use of an interactive computer terminal in the assessment of cognitive function in elderly psychiatric patients. Age and Aging, 1975, 4, 110-115.

Miller, E. A case for Automated Clinical Testing. Bulletin of the British Psychological Society, 1968, 21, 75-78.

Nelson, H. E. A modified card sorting test sensitive to frontal lobe deficits. Cortex, 1976, 12, 313-324.

Overton, G. W. & Scott, K. G. Automated and Manual Intelligence Testing: data on parallel forms of the Peabody Picture Vocabulary Test
American Journal of Mental Deficiency, 1972, **76**, 639-643.

Raven, J. C. Guide to the Progressive Matrices. London: H.K. Lewis, 1960.

Raven, J. C. Personal correspondence, June 1985

Raven, J. C., Court, J. H. & Court, J. Manual for the Raven's Progressive Matrices and Vocabulary Scales: the Coloured Progressive Matrices. London: H.K. Lewis, 1977

Raven, J. C., Court, J. H., Court, J., & Raven, J. Manual for the Raven's Progressive Matrices and Vocabulary Scales: Standard Progressive Matrices. London: H.K. Lewis, 1978.

Raven, J. C., Court, J. H., & Raven, J. Manual for the Raven's Progressive Matrices and Vocabulary Scales: Standard Progressive Matrices. London: H.K. Lewis, 1983.

Rock, D. L. & Nolen, P. A. Comparison of the Standard and computerized versions of the Raven Coloured Progressive Matrices Test.
Perceptual and Motor Skills, 1982, **54**, 40-42.

- Sattler, J. M. Assessment of Children's Intelligence and Special Abilities - Second Edition. Boston: Allyn and Bacon, Inc., 1982.
- Spearman, C. The abilities of man. New York: Macmillan, 1927.
- Stallard, C. K. Computers and education for exceptional children: emerging applications. Exceptional Children, 1982, 49(2), 102-104.
- Tetzrow, C. F. Making child neuropsychological appraisals appropriate for children: An alternative to downward extension of adult batteries. Clinical Neuropsychology, 1983, 5(3), 136-141.
- Watts, K., Baddeley, A., and Williams, M. Automated tailored testing using Raven's Matrices and the Mill Hill Vocabulary tests: a comparison with manual administration. International Journal of Man-Machine Studies, 1982, 17, 331-344.
- Wechsler, D. Wechsler Adult Intelligence Scale - Revised Manual. New York: Psychological Corporation, 1981.
- Weinman, J. A. Detailed computer analysis of performance on a single psychological test. International Journal of Man-Machine Studies, 1982, 17, 321-330.

- Weiss, D. J. (Ed.) Proceedings of the 1979 Computerized Adaptive Testing Conference (Wayzata, Minnesota, June 27-30, 1979)
Minnesota University, 1979 (ERIC Document Reproduction Service No ED 198 164)

VITA

Michael J. Buxton

Birthdate: April 15, 1953

Birthplace: Buffalo, New York

Education:

1981-85 The College of William and Mary in Virginia
Williamsburg, Virginia 23185
Doctor of Education Degree in
Counseling/School Psychology

1975-77 Syracuse University
Syracuse, New York 13210
Master of Science Degree in Counseling

1971-75 Canisius College
Buffalo, New York 14130
Bachelor of Science in Biology - Psychology

Current Position:

School Psychologist
Virginia Beach City Public Schools
Virginia Beach, Virginia 23456

ABSTRACT

THE DEVELOPMENT OF A COMPUTER ASSISTED ASSESSMENT DEVICE FOR USE IN SCREENING CHILDREN REFERRED FOR PSYCHOEDUCATIONAL EVALUATIONS.

Michael J. Buxton, Ed.D

The College of William and Mary in Virginia, July 1985.

Chairperson Roger Ries, Ph.D

Recent advances in microcomputers allowed for the development of a Computer Assisted Assessment (CAA) device which provided a higher degree of feedback and interaction for subjects than had previously been possible. This enhanced human/machine interactive capability was found to increase the comparability between a CAA device and a traditional test instrument. A counter-balanced, pre-post test comparison between the manual and CAA Raven Standard Progressive Matrices (RSPM) indicated no difference between test mediums, with the exception of the CAA pre-test requiring significantly longer for students to complete than the manual pre-test version. A reliability study of the CAA RSPM indicated this version of the RSPM produced a test-retest $r = .88$ with a two-week interval ($N = 21$). In addition to examining the comparability between test mediums, five CAA derived measures were developed to determine their efficacy as psychometric measures. Three measures of response time and two measures of 'motoric efficiency' or purposefulness were developed. Each measure provided a means of fractionating subjects' responses and recording these responses in milliseconds and measuring motoric output with a resolution of 1/80th of an inch. Eighty-five CAA administrations of the RSPM were obtained using 4th, 5th and 6th grade public school students. Discriminate analysis revealed 3 of the CAA derived psychometric measures produced significant results at the $p < .10$ level. A final post hoc comparison between previously administered WISC-R scores and the CAA RSPM was made with 20 students. Multiple regression and correlation analyses indicated significant relationships with the CAA RSPM and the WISC-R verbal and full scale IQ scales. In addition several CAA derived measures demonstrated significant relationships with the WISC-R verbal, performance and full scale IQ scores.